

Content Adaptation, Personalisation and Fine-grained Retrieval: Applying AI to Support Engagement with and Reuse of Archival Content at Scale

Rasa Bocyte^a and Johan Oomen^b

Netherlands Institute for Sound and Vision, Hilversum, The Netherlands
rbocyte@beeldengeluid.nl, joomen@beeldengeluid.nl


Keywords: Reuse, Video Summarisation, Content Adaptation, Personalisation, Multimedia Annotation, Retrieval.


Abstract: Recent technological advances in the distribution of audiovisual content have opened up many opportunities for media archives to fulfil their outward-facing ambitions and easily reach large audiences with their content. This paper reports on the initial results of the ReTV research project that aims to develop novel approaches for the reuse of audiovisual collections. It addresses the reuse of archival collections from three perspectives: content holders (broadcasters and media archives) who want to adapt audiovisual content for distribution on social media, end-users who have switched from linear television to online platforms to consume audiovisual content and creatives in the media industry who seek audiovisual content that could be used in new productions. The paper presents three use cases that demonstrate how AI-based video analysis technologies can facilitate these reuse scenarios through video content adaptation, personalisation and fine-grained retrieval.

1 INTRODUCTION

Audiovisual archives across the globe are looking after massive amounts of digitised and born-digital content. Together, they represent a vital part of our collective memory. Archiving is vitally proactive - it combines acts of selection and preservation with community engagement, storytelling and production (Kaufmann, 2018). Ever changing audience expectations underscore the need for organisations to define a comprehensive access strategy, including an assessment of channels used to research out to their users. Recent advances in the distribution of audiovisual content have opened up many opportunities for archives to fulfil these outward-facing ambitions. Syndicating content over various platforms, using artificial intelligence to extract knowledge and tracking online use to better understand audiences are just a few examples of innovations that will help organisations to maximise their impact. Designing resilient, inclusive, outward-facing audiovisual archives requires experimentation and collaboration between memory institutions, multiple scientific disciplines and the creative industries.

ReTV (<https://retv-project.eu/>) is a pan-European research action that brings computer scientists, broadcasters, interactive TV companies, audio-visual archives from across Europe. It develops and evaluates technology that provides deep insights into the interactions with digital content and re-purposing of media across channels. The resulting Trans-Vector Platform (TVP) built by ReTV empowers content holders (such as broadcasters and archives) to continuously measure and predict the success of content in terms of reach and audience engagement across distribution channels (vectors), allowing them to optimise decision making processes. Specifically for archives, the TVP supports novel approaches for reuse of their collections using video content adaptation, personalisation and fine-grained retrieval. This paper presents the use cases that are central to the work in ReTV and demonstrate how AI-based video analysis technologies could be employed by broadcasters and media archives to support the reuse and repurposing of their audiovisual collections.

^a  <https://orcid.org/0000-0003-1254-8869>

^b  <https://orcid.org/0000-0003-1750-6801>

2 REUSE OF AUDIOVISUAL CONTENT: THREE PERSPECTIVES

The emergence of digital technologies has radically changed the way how audiovisual content is being created, distributed and consumed. Video-on-Demand (VOD), Web-TV and social media platforms are increasingly becoming the primary source for access to audiovisual content and are introducing new forms of media consumption that are not dictated by TV guide schedules or limited to passive consumption from one’s living room (Mikos, 2016). VOD platforms embrace and encourage binge-watching behaviour where audiences spend prolonged time watching consecutive episodes of a television programme of their choice and continuously receive recommendations for more similar content (Jenner, 2017). Social media platforms are characterized by mechanisms for liking, sharing, contributing and (co-)creating that are meant to stimulate high-level of interaction with content and with other users (Dolan, Conduit, Fahy, Goodman, 2016). These new consumer behaviours demand audiovisual content that caters to niche audiences defined by their individual interests and viewing habits.

This changing media landscape creates a market opportunity for audiovisual archives. Audiovisual archives look after a wealth of content that is primarily used by professional media professionals and researchers but it is not easily accessible to the general public. While many cultural organisations create online portals for their online collections (*EUscreen*, *Europeana*), additional outward-facing efforts are needed to ensure that this content reaches audiences who might not know about its existence. By capitalising on their digitised and born-digital collections and making them more accessible, cultural organisations can fulfill consumer demand for content online and bring unprecedented visibility to their audiovisual holdings.

ReTV is developing three use cases that showcase possibilities for the reuse of archival collections in digital contexts from three perspectives:

- *video Content Adaptation and Publication on Social Media* for content owners (broadcasters and media archives) who want to distribute their content with high impact;
- *Personalised Video Delivery* for end-users who have switched from linear television to online platforms and personal devices to consume

audiovisual content that fits their personal interests;

- *Generous Interfaces for Retrieval of Video Segments* for creatives in the media industry who seek audiovisual content that could be reused in new productions.

2.1 Concept: ReTV Trans-Vector Platform

To support various uses for content repurposing and reuse, ReTV is building a system of modular technical components for ingestion, analysis and enrichment of audiovisual content and various data sources. All the components communicate via APIs and can be used in different configurations to support various use cases and build different frontend applications for content repurposing. The backbone of TVP is a module for AI-based video analysis that includes segmentation of video content into scenes, shots and subshots, annotation with semantic concepts, brand detection and video summarisation. These modules are supported by other TVP components, including crawling of online news sources (websites, social media), metadata analysis, temporal annotation, and content recommendation and scheduling services. Together, they ensure that repurposed archival content reaches audiences on appropriate channels at optimal time.

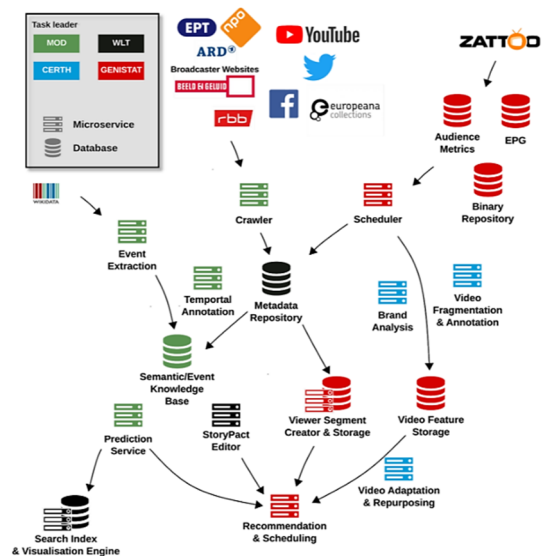


Figure 1: TVP architecture as a modular design of microservices. It indicates the division of tasks between the research partners building the TVP: MOD (Modul University), WLT (WebLyzard), CERTH (Centre for Research and Technology Hellas) and Genstat.

This paper will focus on the AI-based video analysis components within the TVP that facilitate archival content reuse in the three ReTV use cases.

3 USE CASES

The following section outlines three uses cases that make use of AI-based video analysis technologies to repurpose archival audiovisual content - video content adaptation, personalised video content delivery and retrieval of video segments through generous interfaces. Together, they highlight how recent advances in AI can be leveraged by audiovisual archives to support new modes of reuse.

3.1 Video Content Adaptation for Online Publication

The consumption of linear television broadcasting is radically different from the way television content is consumed on social media. Instead of watching full-length programmes, audiences on social media platforms are used to watching short, often muted videos with intertitles, that allow them to quickly survey the essence of the story. This dictates that the format of archival broadcaster content needs to be adapted before publication online if it is to have high impact on online audiences. Since different social media channels have different requirements in terms of optimal length and format, such adaptation currently requires good understanding of the content and lengthy manual editing process to select video segments that attract viewer attention.

For this purpose, ReTV developed a tool that automatically summarises full-length videos into short clips that convey the narrative of the entire video. The tool is built around video analysis and summarisation services that shorten a video into a selection of shots that portray key moments in the story (Apostolidis, Metsai, Adamantidou, Mezaris, Patras, 2019). Since summarisation does not take into account audio elements (the selected shots might be cut in the middle of the sentence), the summarised videos are muted. The tool also provides creative editing functions, allowing users to add overlaying images, text, audio or subtitle track as well as edit the sequence of shots in the video.

To perform the first round of evaluations with professionals from media archives and broadcasting organisations, the tool was tested with a selected number of archival newsreel content from the Netherlands Institute for Sound and Vision collection (accessible via <https://openbeelden.nl/en>). All videos

were between 2-5 minutes long and were summarised into 20-30 second clips.

Users indicated that automatic summarisation would significantly reduce the efforts needed to manually edit videos before publishing them online and would encourage their organisations to share more content on social media. The evaluation results imply that although the summarised videos accurately conveyed the narrative of the original video, the loss of audio track was seen as a negative trade-off. Testers suggested that audio analysis could be performed to provide suggestions for overlaying text and subtitles as well as descriptions accompanying videos. This is particularly pertinent in cases where subtitles are not available. Testers also expressed that they would like to manually control certain editor parameters that would determine the outcomes of video summarisation (e.g. adjust the length of shots in the summary).

Our future work will focus on further adapting video summaries to suit various content genres, publication channels and various audiences, e.g. creating different length summaries for different social media platforms, making different versions of the same video that target different audiences, perform summarisation for multiple videos. Additionally, ReTV will introduce additional components for audio analysis and text editing that would complement visual analysis services. To further evaluate the quality of video summarisations, ReTV will perform tests with consumers and monitor their engagement with summarised content.

3.2 Personalised Video Delivery

Further building on the idea that AI can adapt broadcaster collections to different online publication channels, the second use case explores how audiovisual content could be customised to a single person. The concept of personalising user experience is already established in the broadcasting and media industries - over-the-top (OTT) platforms like Netflix, video-on-demand and streaming platforms like YouTube have adopted systems that track viewing patterns and match them with available content to make content recommendations for each individual user, in this way keeping users engaged for prolonged time and returning to consume more content (Covington, Adams, Sargin, 2016; Lund, Ng, 2018). It is harder for media archives to achieve the same effect since their online collection portals are less concerned with entertainment and more with presenting digital collections in a contextualised, informative and educational form. Therefore, to

benefit from personalisation, media archives need alternative channels to distribute their content.

This use case investigates how archives can personalise their audiovisual collections using AI-driven video analysis and distribute content to audiences using digital communication channels. The fine-grained video analysis creates affordances for a more precise matching between user's profile and video content - retrieving relevant content for the user based on scene- and shot-level analysis rather than recommending content based on semantic relationships between video titles. This opens possibilities to match individual segments of archival content to particular users. Additionally, AI-driven video summarisation can be employed to give users an initial overview of recommended content and help them decide whether they are interested in watching it.

To determine what kind of personalised content users would like to receive from a media archive and on what criteria such personalisation should be based, ReTV developed a chatbot that delivers archival video content. Chatbot was selected as the most appropriate channel for distribution since it allows us to collect accurate data for building personal profiles. A mixture of implicit and explicit feedback is seen as the most appropriate method to set up user profiles, i.e. monitoring viewing habits as well as asking explicit questions about personal preferences (Hobson and Kompatsiaris, 2006). The bi-directional communication with users through the chatbot allows us to obtain immediate implicit feedback as well as monitor viewing patterns.

Once users sign up to use the chatbot, they are immediately directed to set up their profile where they can choose topics they are interested in. For this first iteration of the ReTV chatbot, a small set of videos from the Netherlands Institute for Sound and Vision collection was selected and grouped under five high-level categories - Sports, Animals, Fashion, Dutch Life and Culture, History. These high-level categories are similar to what users would find when browsing through archival collections online; here, they serve as a starting point for understanding what kind of categorisation - if any - would support content personalisation for individual users.

The chatbot delivers content that corresponds to the user's selected topics on a regular basis. First, users receive a 1-2 second GIF preview of a video which is created using the video summarisation service - here the GIF summary serves as a teaser to enable users to choose whether they are interested in the content. Next, users can choose to watch the extended summary, see the full-length video or go to

see the item in the online collection operated by the archive.

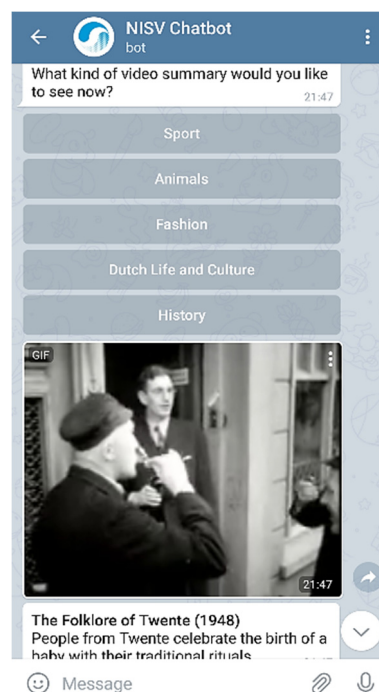


Figure 2: Screenshot of the ReTV chatbot.

Results from evaluation with consumers indicated that users are mostly interested in receiving archival content personalised to their particular context - e.g. news stories about their hometown, historic footage of their favourite football team, programmes broadcasted on this day in history. High-level categorisation of content does not accommodate such personalisation as user interests defy strict boundaries and require much more fluid categorisation based on granular annotations. This suggests that the next iteration of the chatbot should focus on giving users more room to indicate their specific interests, instead of limiting their choices to a number of predefined categories dictated by the available content. AI-driven video analysis would then be used to identify video segments that correspond to those interests and compile those segments into one video. This method would provide more accurate personalisation for the users and would help media archives utilise media assets that might not be relevant for larger audiences.

One of the challenges for the future research is to find a way to personalise content in a way that allows users to discover new things rather than consuming content about topics they are already familiar with.

3.3 Generous Interfaces for Retrieval of Video Segments

The final use cases approaches reuse of broadcaster collections from the perspective of creatives working in the media industry - producers, journalists, documentary makers, amateur video content makers creators, etc. The most common point of interaction between these creators and archival collections is the Media Asset Management (MAM) systems. The interface of these systems is often closed off - the content is hidden behind the search bar which means that the discovery of content is highly depended on one's tacit knowledge. In addition, MAMs provide access to the full item whereas creators are interested in much more fine-grained access that allows them to identify and retrieve specific segments of content (Huurink, Hollink, Van Den Heuvel, De Rijke 2010). Lack of granular metadata also means that the search is often restrained by item titles and high-level descriptions. To address these shortcomings, fine-grained annotations created with AI can be employed to facilitate the discovery of relevant content for repurposing in new productions.

This third use case builds on the concept of generous interfaces - harnessing data visualisations to showcase the abundance and richness of digital collections (Whitelaw, 2015). Many cultural heritage organisations have adopted this for their online collections as a way to help users navigate through a large number of items available and find often unexpected links between them. Equally, generous interfaces work on an individual item level where they provide various facets to inspect the digital object (e.g. providing a breakdown of a colour scheme in a painting).

While most of the current generous interfaces work with image-based collections, the application of this concept to time-based media would be exceedingly beneficial as it would allow users to retrieve, inspect and compare different segments of the same item or a number of items at a glance.

The foundation for ReTV's generous interface is the visual analysis service. Video segmentation into individual shots allows us to provide a breakdown of all concepts detected on a shot level. It is also possible to perform concept-based search and retrieve all shots from a video that contain a specific concept. Such fine-grained functions increase the possibilities of locating more unexpected video segments, especially for content that would not be discovered otherwise due to the lack of descriptive metadata.



Figure 3: Interface of the visual analysis service with scene and shot segmentation.

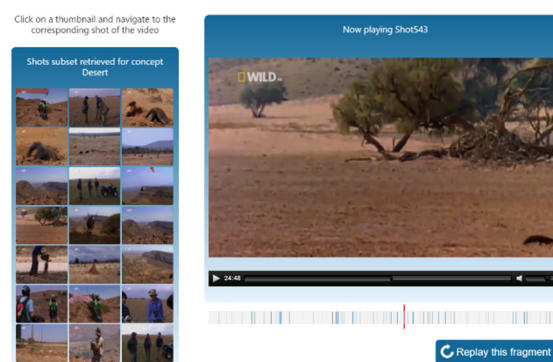


Figure 4: Interface of the visual analysis service showing the retrieval of shots that depict a selected concept.

For creators using MAMs, a generous interface based on visual analysis of time-based media could provide significant support for conducting “image search” - finding relevant video segments that could be reused to illustrate narratives in their productions. It creates affordances for a more serendipitous content retrieval, one that can guide and inspire new creative narratives (Sauer, 2017). In particular, it would make archival systems more accessible to amateur creators who do not have pre-existing knowledge about broadcaster collections. To provide an even more comprehensive analysis of audiovisual content, such generous interfaces could combine a number of AI-powered services, e.g. concept detection together with face recognition and speech analysis.

4 CONCLUSIONS

The initial evaluation of the three ReTV use cases presented in this paper points to the wide-reaching impact that reuse and distribution of archival

collections can have on media archives and various user groups who engage with audiovisual heritage. ReTV is a three year project and has just past its mid-term review. In the next period, and as the TVP is developed further, more user testing will be conducted. Also, active consultations are planned with stakeholders from the heritage sector and creative industries. Both the user testing and the consultations inform the further development and eventual deployment of the final system.

With automated video content adaptation and publication on social media platforms, media archives can highlight the value of audiovisual heritage in contemporary contexts, specifically related to social media. Delivery of customised videos help users to develop more personalised and long-lasting relationships with archival collections. And finally, the fine-grained access to the archive through generous interfaces welcomes professional and amateur creators to discover the creative potential of audiovisual heritage.

ACKNOWLEDGEMENTS

This work was supported by the EUs Horizon 2020 research and innovation programme under grant agreement H2020-780656 ReTV.

REFERENCES

- Apostolidis, E., Metsai, A., Adamantidou, E., Mezaris, V., & Patras, I., 2019. A Stepwise, Label-based Approach for Improving the Adversarial Training in Unsupervised Video Summarization. *1st Int. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV'19) at ACM Multimedia 2019*. <http://doi.org/10.1145/3347449.3357482>
- Covington, P., Adams, J., & Sargin, E., 2016. Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198. <https://doi.org/10.1145/2959100.2959190>
- Dolan, R., Conduit, J., Fahy, J., & Goodman, S., 2016. Social Media Engagement Behaviour: a Uses and Gratifications Perspective. *Journal of Strategic Marketing*, 24(3-4), 261–277. <https://doi.org/10.1080/0965254X.2015.1095222>
- Hobson, P., & Kompatsiaris, Y., 2006. Advances in semantic multimedia analysis for personalised content access. *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*. <https://doi.org/10.1109/ISCAS.2006.1693027>
- Huurnink, B., Hollink, L., Van Den Heuvel, W., & De Rijke, M., 2010. Search behavior of media professionals at an audiovisual archive: A transaction log analysis. *Journal of the American Society for Information Science and Technology*, 61(6), 1180–1197. <https://doi.org/10.1002/asi.21327>
- Jenner, M., 2017. Binge-watching: Video-on-demand, quality TV and mainstreaming fandom. *International Journal of Cultural Studies*, 20(3), 304–320. <https://doi.org/10.1177/1367877915606485>
- Kaufman, P. B., 2018. *Towards a New Audiovisual Think Tank for Audiovisual Archivists and Cultural Heritage Professionals*. Netherlands Institute for Sound and Vision. <https://doi.org/10.18146/2018thinktank01>
- Lund, J., & Ng, Y., 2018. Movie Recommendations Using the Deep Learning Approach. *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, 47–54. <https://doi.org/10.1109/IRI.2018.00015>
- Mikos, L., 2016. Digital Media Platforms and the Use of TV Content: Binge Watching and Video-on-Demand in Germany. *Media and Communication*, 4(3), 154–161. <https://doi.org/10.17645/mac.v4i3.542>
- Sauer, S., 2017. Audiovisual Narrative Creation and Creative Retrieval: How Searching for a Story Shapes the Story. *Journal of Science and Technology of The Arts*, 9(2), 37–46. <https://doi.org/10.7559/citarj.v9i2.241>
- Whitelaw, M., 2015. Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly*, 9(1). <http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html>