2021 IEEE Int. Symp. on Multimedia (ISM). Author's accepted version. The final publication is available at https://ieeexplore.ieee.org/

A Web Service for Video Smart-Cropping

Konstantinos Apostolidis, Vasileios Mezaris
CERTH-ITI
Thessaloniki, Greece, 57001
kapost@iti.gr, bmezaris@iti.gr

***Abstract*—This paper presents a Web service that supports the automatic transformation of a video's aspect ratio. We employ a modified smart-cropping technique from the literature that aims to minimize the loss of semantically important visual content. We integrate this method in an easy-to-use publicly-accessible Web service where a video can be uploaded and automatically transformed to the desired aspect ratio. We also demonstrate that the algorithmic modifications we introduced in the process of building our Web service offer performance gains, when compared to the original method of the literature.**

***Keywords*-video aspect ratio transformation; smart video cropping; web service**

I. Introduction

Videos created for traditional TV and desktop computer monitors are typically consumed in landscape aspect ratios (e.g., 16:9 or 4:3). However, on more modern and nowadays ubiquitous mobile devices (mobile phones and tablets) video content is optimally displayed in different aspect ratios. Additionally, people record videos in portrait mode on their smartphones and these may need to be cropped to landscape mode before being published in online video sharing platforms, which now dictate the use of specific video aspect ratios. Existing videos, or summaries of them targeting different channels and audiences [1], would have to be transformed to be distributed and optimally displayed on these platforms and mobile devices. A straightforward approach for transforming a video to a different aspect ratio would involve either static cropping of content or padding the frames with black borders to reach the target aspect ratio. However, such simple approaches produce unsatisfactory results. On the other hand, common video aspect ratio transformation methods of the literature, such as warping [2] and seam carving, often introduce distortions and may alter the semantics of the video. In practice, this makes the generation of tailored versions of the video content a tedious manual or semi-automatic task. For an analysis of the literature on video retargeting methods the interested reader is referred to [3] (aspect ratio transformation) and [1] (deep-learning-based summarization).

While several research works and some commercial software for transforming a video's aspect ratio are available, there is a lack of free video retargeting tools that can be used by anyone without requiring any experience. To remedy this, we built a freely accessible Web application that enables users to submit locally-stored or on-line available videos and automatically transform them to a desired target aspect ratio.

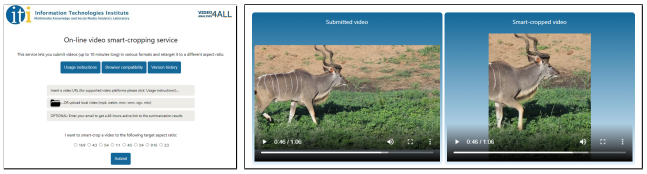
II. Video Smart-Cropping Web Service

We designed a framework that consists of: a) a REST service, which hosts the developed technologies for video cropping (backend) and b) an interactive user interface (UI) that allows the user to exploit the functionality of our Web service. In particular through the UI of this tool, the user is able to: a) submit a video for analysis (either available on-line or locally stored in the user's device), b) select the target aspect ratio from a list of predefined aspect ratios, and c) get the transformed video in a way that enables both immediate on-line inspection through the UI of our tool and the downloading of the video file in the user's device.

A. Method and backend

In [3], a smart-cropping technique is proposed that aims to minimize the loss of semantically important visual content. This method infers the viewer's attention by computing the saliency map for each frame, employing a DCNN-based saliency detection method [4]. The saliency may be concentrated in a small region of the whole frame or be in the form of multiple blobs. Aiming to select the main part of the viewer's focus, [3] employs a filtering-through-clustering procedure, where the location and value of the non-zeroed items of the saliency map are clustered and the cluster with the highest weight is selected, discarding the rest of the items. The center of mass of the selected cluster is considered as a single point center of the viewer's attention. A crop window for each frame is computed based on the displacement of the center of mass. Finally, a number of techniques are employed to smooth the movement of the crop window.

We introduce a variation of [3] featuring the use of an optimized set of parameters as well as three modifications. First, to speed-up the retargeting process we spatially sub-sample the saliency maps that are fed into the filtering-through-clustering procedure by a factor of 8. Second, to improve the quality of the final video, we devise a simple "focus stability" mechanism. Our simple intuition behind this mechanism is that when the center of the crop window moves over areas of low saliency on the saliency map, this



(a) (b)

Figure 1. a) The landing page of the frontend, b) The results page of the frontend and an example result of transformation from 16:9 to 4:5.

signifies a shift of focus. If two subsequent opposite shifts of focus are temporally closeby (i.e., less than one second apart) we reject them, in order to avoid sudden changes of focus which may result in visually displeasing crop window movements. Finally, instead of the LOESS method for time-series smoothing we utilize a Savitzky-Golay filter, since visual inspection of the results for both approaches revealed that the latter yields a smoother crop window movement.

Our modified smart-cropping method is deployed as a REST service that: a) retrieves a video file, b) analyzes the video using the above-described modified version of the method of [3], c) transforms the video frames to the target aspect ratio, and d) renders the transformed video.

B. User interface

Similarly to previously-developed Web services, such as [5] for video summarization, we have designed an interactive user interface (see Fig. 1a) which allows the user to submit videos and transform their aspect ratio. Video submission is performed on a one-by-one basis (i.e., no video collection analysis is supported) and, for demonstration purposes, the submitted videos are allowed to be up to 10 minutes long and 200MB in file size.

In particular, to submit a video for retargeting the user can either provide its URL or upload a local copy of it from his/her device. The supported on-line video sources include YouTube, Facebook, Twitter, Instagram, Vimeo, DailyMotion, LiveLeak and Dropbox. The service can handle videos in mp4, webm, avi, mov, wmv, ogv, mpg, flv, and mkv formats. After submitting a video, the user can monitor the progress of the retargeting procedure, and is also able to submit additional requests while the previous ones are being analyzed. The submitted and the transformed video are cached in a server for 24 hours, and after this time period, both videos as well as the user's e-mail address (if provided) are automatically deleted from the server.

When the analysis is completed, an automatic refresh of the user interface is performed and the transformed video is presented to the user (see Fig. 1b). Optionally, if the user provided an e-mail address she/he may close the Web browser and be notified by e-mail when the transformed video is ready. Furthermore, the user is able to download the transformed video.

III. Deployment and Experiments

Our modified smart-cropping method is evaluated on the RetargetVid [3] video retargeting benchmark dataset. Specifically, we use the evaluation protocol and ground-truth croppings of [3] to compare the original method [3] and the one used in our service. The comparison results in Table I show that our algorithmic modifications result in the averaged Intersection-over-Union (IoU) scores for all video frames being increased by approx. 3 percentage points, as well as in a 30% speed-up. For details on the evaluation protocol, metrics used, and more comparisons, the interested reader is directed to [3].

Table I
Comparison of the method of [3], and the modified method we introduced in our Web service

Method	Worst	Best	Mean	t (%)
Results (IoU, %) for 1:3 target aspect ratio				
[3]	48.6	50.9	49.9	19
[3] + our modifications	**51.7**	**53.8**	**52.9**	**13**
Results (IoU, %) for 3:1 target aspect ratio				
[3]	70.1	73.6	71.4	20
[3] + our modifications	**74.4**	**77.0**	**75.3**	**14**

Our Web service for video retargeting can be accessed and tested at multimedia2.iti.gr/videosmartcropping/service/start.html, while a short instructional video can be found at https://youtu.be/_pdTDMWbIfs.

Acknowledgment

This work was supported by the EU's Horizon 2020 R&I programme under grant agreement H2020-780656 ReTV.

References

[1] E. Apostolidis, E. Adamantidou, A. Metsai, V. Mezaris, and I. Patras, "Video summarization using deep neural networks: A survey," *Proceedings of the IEEE*, accepted for publication, 2021.

[2] H. Nam, D. Park, and K. Jeon, "Jitter-robust video retargeting with kalman filter and attention saliency fusion network," in *Proc. EEE Int. Conf. on Image Processing (ICIP)*, 2020, pp. 858–862.

[3] K. Apostolidis and V. Mezaris, "A fast smart-cropping method and dataset for video retargeting," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2021, pp. 1956–1960.

[4] R. Droste, J. Jiao, and J. A. Noble, "Unified image and video saliency Modeling," in *Proc. 16th European Conf. on Computer Vision (ECCV)*, 2020.

[5] C. Collyda, K. Apostolidis, E. Apostolidis, E. Adamantidou, A. I. Metsai, and V. Mezaris, "A web service for video summarization," in *ACM Int. Conf. on Interactive Media Experiences (IMX)*, 2020, pp. 148–153.