COZI: Crowdsourced and Content-based Zoomable Video Player

Axel Carlier
IRIT-ENSEEIHT
University of Toulouse
carlier.axel@gmail.com

Arash Shafiei
IRIT-ENSEEIHT
University of Toulouse
arash.shafiei@gmail.com

Julien Badie
IRIT-ENSEEIHT
University of Toulouse
julien.badie@gmail.com

Salim Bensiali
IRIT-ENSEEIHT
University of Toulouse
salim.bensiali@gmail.com

Wei Tsang Ooi
Dept. of Computer Science
Nat. Univ. of Singapore
ooiwt@comp.nus.edu.sg

ABSTRACT

We present a new user interface designed to allow easy yet effective zooming and panning into high-definition videos for playback on low resolution displays. Our system first applies state-of-the-art video analysis algorithms to detect salient regions of interest and recommends them to users. These recommendations help users to quickly identify important regions in the video and zoom into the regions with a single mouse click. The salient regions may moves according to the movement of track objects, further reducing the need for users to manually pan to track an object of interests. To further improve the relevance of the recommended regions, users' interactions are logged and analyzed. The actual regions selected and viewed by users serve as a feedback and is integrated into the system to improve the recommendations. We have implemented a Web-based version of the user interface, running on modern browsers supporting HTML5. We describe the algorithms and optimizations used to implement and improve the system.

Categories and Subject Descriptors: H.5.1 [Multimedia Information Systems]: Video

General Terms: Human Factors, Design

Keywords: Interaction Techniques, Zoomable Video, Content Analysis, Crowdsourcing

1. INTRODUCTION

Resolution of videos are getting higher and higher (up to $7,680 \times 4,320$ pixels) while the display size of electronic devices such as laptops or smartphones remains limited. This observation motivated the study of new paradigms of interaction such as zoomable video, first presented in [2]. In this work, Ngo et al. propose different compression algorithms to adapt to this new form of interaction and present an interface that includes functionalities such as zooming in and out the video, and moving the selected viewport by dragging with the mouse (panning).

Although the work by Ngo et al. provides some insight into the ways users interact with zoomable video (studied in [1]), this interface can be tedious and sometimes confusing to use. Indeed, when zooming in a moving object, a user might be disoriented because the object leaves the viewport. The action of panning needed to follow the object is not very intuitive, and must sometimes be repeated several times in order to track the object along its entire movement.

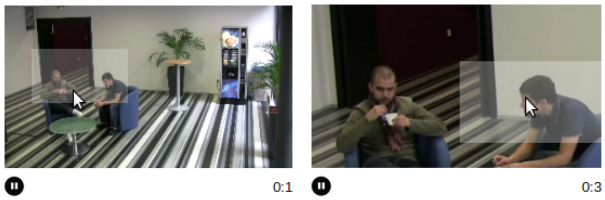


Figure 1: Our zoomable video interface

Motivated by that problem, we decided to develop a new zoomable video interface in which we recommend viewports to users. We call our new interface *COZI*, short for COntent-aware Zoomable Interface. Figure 1 shows an example of such a recommendation in COZI: when the mouse cursor hovers over a region of interest (ROI), i.e., a region we detected as potentially interesting, the recommended viewport appear as a semi-transparent white rectangle. After left clicking inside the recommended viewport, the user zooms in and views the corresponding ROI at a higher resolution. The detected ROIs might be moving. In that case, the recommended viewport automatically tracks the object, which considerably reduces the number of interactions needed from the user. The user may left click outside the recommended viewports. In which case, the interface would zoom in by one zoom level, centered around the click position. Users can right click to zoom out by one zoom level. As in the interface proposed by Ngo et al., users can drag with a mouse to pan while zoomed into the video.

In this paper, we describe the architecture of our web-based zoomable video interface, as well as the content analysis algorithms we implemented to detect ROIs and generate the recommended viewports. We also show how users' interactions can be added as an input to improve the quality of the recommendations.

2. SYSTEM IMPLEMENTATION

The overview of COZI is presented in Figure 2. There are two parts for this system: an off-line content analysis process and a web application.

The web application was developed using new features of HTML5 such as canvas and video tags. We handle users' interactions with Javascript, allowing us to collect those interactions into

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.
ACM 978-1-4503-0616-4/11/11.

a remote MySQL database. The interface takes a JSON (JavaScrip Object Notation) file as an input, which is downloaded along with the video. This JSON file is generated during a video analysis offline process and contains the information (time, position and size) of the recommended viewports.

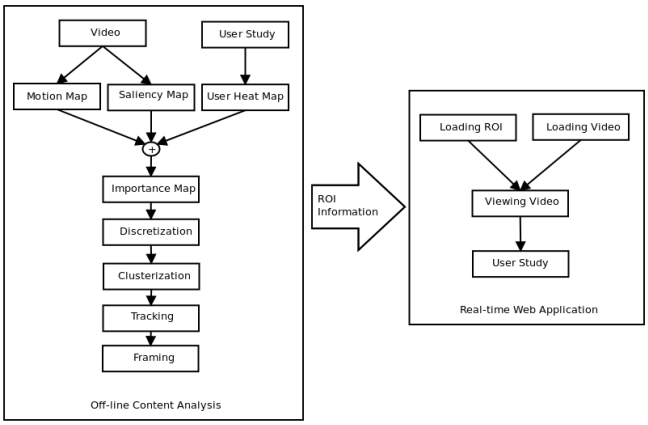


Figure 2: System Overview

2.1 Importance Map Generation

We now describe the three main steps of our content analysis. The first step to detect ROIs on the video is to apply some content analysis algorithms. We compute a saliency map (Figure 3(b)) based on color gradients and strong luminosity variations. Motion is also detected by comparing color maps from successive frames (Figure 3(a)). Finally, we generate an importance maps by summing and normalizing the saliency and motion maps (Figure 3(c)). Brightest regions of this importance maps are likely to be ROIs, and therefore they are recommended as viewports in our interface.

(a) Motion (b) Saliency (c) Importance

Figure 3: (a) Motion map (b) Saliency map (c) Importance map

2.2 Viewport Optimization

We now briefly describe how we get viewports out of this importance map. We first discretize the importance map by generating a set of points following the probability distribution described by the importance map. We then clusterize this set of points using a Mean-shift algorithm. At this point we have clusters of points in every frame. Our next step is to track clusters over time, by looking for the closest clusters in successive frames.

The last step, which we call *framing*, consists in positioning optimally the viewports on the frames. We allow in our interface three different levels of zooming, which means the viewports can be of three different sizes (see Figure 4). Framing is therefore an optimization algorithm where we aim at placing viewports at the best position regarding several criterias, such as maximizing the number of points contained in a viewport, or minimizing the cuts of clusters. Figure 4 shows the output of our framing algorithm. We can see that in this example the viewports are focused on the different characters of the scene.

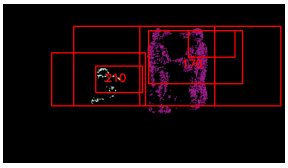
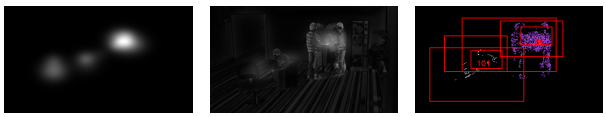


Figure 4: Output of our *Framing* algorithm

2.3 Combining User Feedback with Content Analysis

In the following section we describe how we take users' interations in account to recommend better viewports.

On every frame, we know which viewports have been visualized by the users. We decide to model the attention of a user as a gaussian centered in the viewport he visualized. By adding the gaussians obtained from every users on a single frame, we compute a *user map*, as shown in Figure 5(a). We then reapply our entire algorithm on a new importance map that includes the user map. The new importance map is shown as Figure 5(b), along with the new framing results (Figure 5(c)). The viewports are now focusing on an object laying on the table between the two characters, which is indeed a ROI of the scene that is neither salient nor moving, making it almost impossible for a classic content analysis algorithm to detect.



(a) User Map (b) New Importance Map (c) New Framing results

Figure 5: (a) User heat map (b) Importance map containing user heat map

3. CONCLUSION

In this demo, we briefly present COZI, a novel interface for zoomable video based both on content analysis and crowdsourcing. This interface provides recommendations to users, by indicating to them the regions that are likely to be interesting. We conducted a user study (we can not present it here because of the lack of space), and results show our interface not only provides better viewing experience with fewer interactions, but also enhances users' understanding of the videos by showing them where to look.

4. REFERENCES

[1] A. Carlier, R. Guntur, and W. T. Ooi. Towards characterizing users' interaction with zoomable video. In *Proceedings of the 2010 ACM workshop on Social, adaptive and personalized multimedia interaction and access*, pages 21–24, Florence, Italy, 2010.

[2] K. Q. M. Ngo, R. Guntur, A. Carlier, and W. T. Ooi. Supporting zoomable video streams via dynamic region-of-interest cropping. In *Proceedings of MMSys'10*, pages 259–270, Scottsdale, AZ, USA, 2010.