A Multi-Video Browser for Endoscopic Videos on Tablets

Marco A. Hudelist
Institute for Information Technology
Klagenfurt University
Klagenfurt, Austria
marco@itec.aau.at

Sabrina Kletz
Institute for Information Technology
Klagenfurt University
Klagenfurt, Austria
sabrina@itec.aau.at

Klaus Schoeffmann
Institute for Information Technology
Klagenfurt University
Klagenfurt, Austria
ks@itec.aau.at

ABSTRACT

We present a browser for endoscopic videos that is designed to easily navigate and compare scenes on a tablet. It utilizes frame stripes of different levels of detail to quickly switch between fast and detailed navigation. Moreover, it uses saliency methods to determine which areas of a given keyframe contain the most information to further improve the visualization of the frame stripes. As scenes with much movement can be non-relevant *out-of-patient* scenes, the tool supports filtering for scenes of low, medium and high motion. The tool can be especially useful for patient debriefings as well as for educational purposes.

CCS Concepts

•**Computer systems organization** → **Embedded systems;** *Redundancy;* Robotics; •**Networks** → Network reliability;

Keywords

Video Browsing; Mobile; Endoscopic Videos; User Interface

1. INTRODUCTION & RELATED WORK

Powerful mobile devices like tablets and smartphones have had a large impact on the medical sector over the last years. For example, they enable doctors to get updates about their patients on-the-go when they are moving from patient to patient and surgery to surgery. Nevertheless, there is still room for improvement. Intuitive and effective inspection of endoscopic videos is challenging, as traditional video players are not optimized for this specific task. They can not support users in navigating such videos in a fast and effortless way. Moreover, if surgeons try to compare two recordings of endoscopic procedures the user experience is very cumbersome, as they have to operate two different players at the same time. On smaller screens it may be even necessary to constantly switch between two player windows or applications all the time.

MM '16 October 15–19, 2016, Amsterdam, Netherlands

DOI: 10.1145/1235

Therefore, we propose a video browsing tool for tablets that improves the experience on two fronts: First, it utilizes the concept of *frame stripes* [14, 11] at different levels of detail to speed up navigation and enable iterative browsing approaches. Users are able to start at a coarse but fast navigation level and gradually refine their search by switching to slower but detailed browsing. Moreover, they can switch between uniform sampled segments and content-based segmentation. Filtering segments is also possible by choosing between four options of detected motion in the content, i.e., to exclude out-of-patient segments.

Second, the tool enables users to open two videos at the same time and compare them side-by-side. This feature can be especially useful for comparing the same procedure of different patients or comparing earlier vs. later conditions during a surgery. Navigation and playback can also be synchronized between both videos, so that users do not need to control playback for each individually.

While video browsing for mobile devices tailored for endoscopic recordings is still a rather new field, there is a lot of related work for general video content. ProPane presented by Ganhör [2] is a smartphone video navigation tool that enables fast and very precise browsing at the same time by using various touch gestures. Huber et al. [3] present with *Wipe'n'Watch* an interface for smartphones that focuses on browsing interrelated eLearning video collections. A hierarchical storyboard is utilized by *HiStory*, a tablet interface shown by Hürst and Darzentas [7] that uses a hierarchical storyboard for exploring video content. Hürst et al. also show various other approaches for mobile video browsing like timeline-based interface optimization or even one-handed video navigation for smartphones [8, 9]. Karrer et al. [10] show a system that is using direct manipulation of objects for navigation in a video. Other examples for tablet video browsing interfaces are an interactive 3D filmstrip, a thumb-optimized interface and an interface that uses sub-shot visualization shown by Hudelist et al. [4, 5, 6], a smartphone interface that utilizes a *scrubbing wheel* and a tablet interface where users control playback speed using wipe gestures presented by Schoeffmann et al. [12, 13].

2. INTERFACE

At the top of the interface two playback windows are placed. The two windows can show the same video file at varied positions or two different videos, e.g., to compare the same type of surgery of two different surgeons. A blue border around a window indicates which one will respond to navigational input.

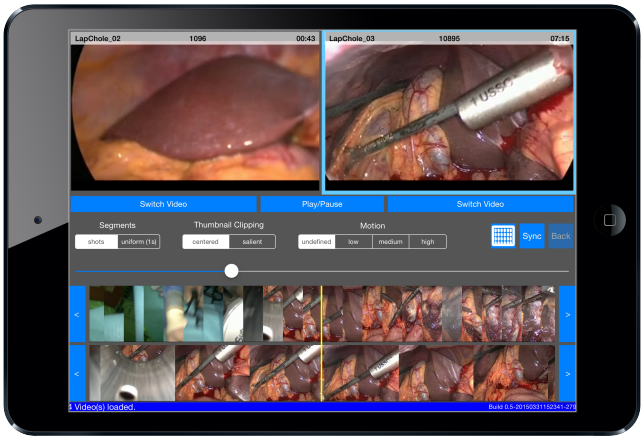

Figure 1: Screenshot of the interface with player windows at top, settings controls in the middle and browsing stripes at the lower half.

At the bottom of the interface two stripes of frames are visible. This visualization is inspired by frame stripes shown by Schoeffmann et al. [14], which are inspired by MO-images, shown by Mueller-Seelich and Tan [11].

Each frame represents a video segment and the stripes can be used to navigate inside a video. For this, users just have to drag the stripes with touch gestures or use the arrow buttons on either sides of the stripes. The stripes are always in sync with each other, so that when users navigate with one of them the other one instantly reorients. As a result the the same content is always below a virtual play head indicated by a yellow line in the middle on the stripes (see Fig. 2).

To offer different navigation approaches the two stripes show the same keyframes but with varying slice sizes. This enables users to traverse between fast but coarse browsing with the top stripe to slower but more detailed browsing with the lower stripe. Moreover, users can always tap on any of the frames to directly navigate to the appropriate position in the video. They can also page through the keyframes by using arrow buttons at each side of the stripes.

In the middle interface section, additional settings of the browsing experience can be changed. On one hand, the type of segmentation can be alternated between shot-based browsing and uniformly sampled segments of one second of length.

To provide users a simple but effective filtering option it is possible to restrict the shown segments according to their contents motion. As *out-of-patient* scenes are often not relevant but contain a high degree of motion due to handling the instrument the setting can be used to focus on more important sections of the video.

Furthermore, three additional buttons let users: (i) switch to an overview browsing mode, (ii) synchronize playback of both video windows, and (iii) undo any earlier performed navigational action like navigating by tapping on a frame. In overview mode the top stripe visualizes the segments by small thumbnails. This enables users to browse a video event faster, as more segments can segments can be inspected at a single glance.

Users can navigate a video also by utilizing a traditional seeker bar in the middle of the screen. This UI control was

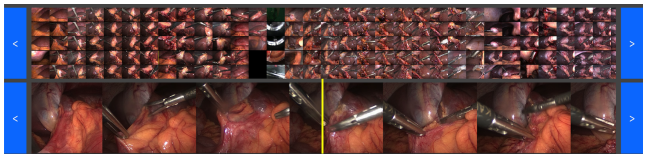

Figure 2: Overview-Mode: Segments are visualized by small thumbnails in the upper stripe for even faster browsing.

added in order to give users an additional interaction element that they know from traditional video players. Moreover, it enables even faster seeking through videos if required without the advantages of instant content insight provided by the navigation stripes. As users interact with it, the stripes re-position themselves accordingly.

2.1 Keyframe Clipping

The way the frames are clipped to the slices can be changed. As it can be important to which area of a given keyframe this clipping is applied users can switch between always clipping to the center of a frame or clipping to the salient region. To detect salient regions in the frame we use the frequency-based method proposed by Achanta et al. [1]. Therefore, we compute the saliency map S for each frame F of size $(W \times H)$ based on the following formula:

$$S(x, y) = |F_\mu - F_G(x, y)| \tag{1}$$

where F_μ is the arithmetic mean pixel value of the image and F_G is a Gaussian blurred version (see [1] for details). In that saliency map we find the horizontal area P_w with a predefined width $W_P \leq W$ that has the maximum average saliency for the complete vertical area:

$$P_w(W_P, H) = \underset{x \in W_P, y \in H}{\arg\max} \frac{\sum S(x, y)}{W_P \times H} \tag{2}$$

3. CONCLUSIONS

We have presented a video browser for tablets that focuses on easy navigation and comparison of different phases of a recorded endoscopic surgery. Two video player windows can be used to navigate freely in two different recordings or to compare different phases of one surgery. For improving navigation it uses two frame stripes with different levels of granularity. One can be used for fast but coarse navigation while the other supports detailed inspection. An additional stripe visualization uses a grid organization of small thumbnails to speed up navigation even more. Moreover, it supports filtering video content for different levels of movement to remove i.e., non-relevant out-of-patient sequences. At our demo booth conference participants will be able to go hands-on with our prototype and test its functionality.

4. ACKNOWLEDGMENTS

This work was supported by Universität Klagenfurt and Lakeside Labs GmbH, Klagenfurt, Austria and funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant KWF 20214 u. 3520/ 26336/38165.

5. REFERENCES

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*, pages 1597–1604. IEEE, 2009.

[2] R. Ganhör. Propane: Fast and precise video browsing on mobile phones. In *Proc. of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, pages 20:1–20:8, New York, NY, USA, 2012. ACM.

[3] J. Huber, J. Steimle, R. Lissermann, S. Olberding, and M. Mühlhäuser. Wipe'n'watch: Spatial interaction techniques for interrelated video collections on mobile devices. In *Proc. of the 24th BCS Interaction Specialist Group Conference*, BCS '10, pages 423–427, Swinton, UK, UK, 2010. British Computer Society.

[4] M. A. Hudelist, K. Schoeffmann, and L. Boeszoermenyi. Mobile video browsing with a 3d filmstrip. In *Proc. of the 3rd ACM Conference on Int. Conf. on Multimedia Retrieval*, ICMR '13, pages 299–300, New York, NY, USA, 2013. ACM.

[5] M. A. Hudelist, K. Schoeffmann, and L. Boeszoermenyi. Mobile video browsing with the thumbbrowser. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 405–406, New York, NY, USA, 2013. ACM.

[6] M. A. Hudelist, K. Schoeffmann, and Q. Xu. *Proc. of the 21st Int. Conf. on MultiMedia Modeling*, chapter Improving Interactive Known-Item Search in Video with the Keyframe Navigation Tree, pages 306–317. Springer, Cham, 2015.

[7] W. Hürst and D. Darzentas. History: A hierarchical storyboard interface design for video browsing on mobile devices. In *Proc. of the 11th Int. Conf. on Mobile and Ubiquitous Multimedia*, MUM '12, pages 17:1–17:4, New York, NY, USA, 2012. ACM.

[8] W. Hürst, G. Götz, and M. Welte. A new interface for video browsing on pdas. In *Proc. of the 9th Int. Conf. on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '07, pages 367–369, New York, NY, USA, 2007. ACM.

[9] W. Hürst and P. Merkle. One-handed mobile video browsing. In *Proc. of the 1st Int. Conf. on Designing Interactive User Experiences for TV and Video*, UXTV '08, pages 169–178, New York, NY, USA, 2008. ACM.

[10] T. Karrer, M. Wittenhagen, and J. Borchers. Pocketdragon: A direct manipulation video navigation interface for mobile devices. In *Proc. of the 11th Int. Conf. on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 47:1–47:3, New York, NY, USA, 2009. ACM.

[11] H. Mueller-Seelich and E. Tan. Visualizing the semantic structure of film and video, 2000.

[12] K. Schoeffmann and L. Burgstaller. Scrubbing wheel: An interaction concept to improve video content navigation on devices with touchscreens. In *IEEE Int. Symposium on Multimedia (ISM)*, pages 351–356, Dec 2015.

[13] K. Schoeffmann, K. Chromik, and L. Boeszoermenyi. Video navigation on tablets with multi-touch gestures. In *IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW)*, pages 1–6, July 2014.

[14] K. Schoeffmann, M. Taschwer, and L. Boeszoermenyi. The video explorer: A tool for navigation and searching within a single video based on fast content analysis. In *Proceedings of the First Annual ACM SIGMM Conference on Multimedia Systems*, MMSys '10, pages 247–258, New York, NY, USA, 2010. ACM.