Dynamic Hierarchical Visualization of Keyframes in Endoscopic Video

Jakub Lokoč*, Klaus Schoeffmann and Manfred del Fabro**

*SIRET Research Group, Department of Software Engineering, Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic
*Klagenfurt University, Universitaetsstr. 65-67, 9020 Klagenfurt, Austria
lokoc@ksi.mff.cuni.cz,ks@itec.aau.at,manfred@itec.aau.at

Abstract. The after-inspection of endoscopic surgeries can be a tedious and time consuming task. Physicians have to search for important segments in the video recording of an intervention, which may have a duration of several hours. Automatically selected keyframes can support physicians in this task. The problem is that either too few keyframes are selected, missing some important information, or too many keyframes are selected, which overwhelms the user. Furthermore, keyframes of endoscopic videos typically show highly similar content. It is hence difficult to keep track of the temporal context of selected keyframes if they are presented in a grid view. To overcome these limitations, we present a dynamic hierarchical browsing technique for large sets of keyframes that preserves the temporal context in the visualization of the frames.

1 Introduction

In the field of medical endoscopy over the last years clinicians have adopted to archive recordings from endoscopic interventions. The reasons for this development are manifold. First, the videos show a first-hand perspective of the surgeons' work and are hence a good source of information for future interventions of the same patient. Secondly, it can be used as explanatory material for training of young surgeons. Moreover, the recording and archival of endoscopic interventions are enforced by law in some countries (e.g., The Netherlands)[3].

Since an archive of endoscopic video is typically growing quickly, with lots of new and long videos added on a daily basis in a hospital, it is especially important to provide expressive and distinctive keyframes to facilitate later browsing of the video material [5, 1]. However, unlike classical broadcast videos the endoscopic videos are far more complex data for automatic content-based analysis, which makes the selection of a good set of representative keyframes a challenging problem. The content does not contain any shot boundaries, but segments with the same content over a long time period, as well as re-appearing content. Therefore, keyframe extraction with uniform sampling or common keyframe selection strategies [4] typically produce a large result set with hundreds of keyframes showing highly similar content. Browsing such a large result set with a lot of redundancies is an inconvenient process for clinicians who want to get a quick

overview of the most important segments in the video. Moreover, as selected keyframes might be non-linearly distributed along the timeline, the physicians can also lose the notion of the corresponding time location, which makes the orientation in the result set even more complicated.

Therefore, we propose a dynamic hierarchical visualization of keyframes for browsing endoscopic video at different levels of detail. More precisely, our approach first samples keyframes based on motion-flow [2] in the endoscopic video content and then performs a hierarchical clustering with a uniformity constraint and a special visualization that shows where the keyframes are located in the video. The visualization uses a compact layout and starts at the root level of the clustering tree but allows the user to zoom in and out at any temporal position in the video. To optimally preserve the browsing context, results of navigation actions (i.e., appearance of new and disappearance of old keyframes) are smoothly animated, which allows for convenient browsing of endoscopic video even for longer recordings with hundreds of keyframes.

2 Keyframe Extraction and Hierarchical Visualization

In order to present potentially highly ranked representatives in a user-friendly way and keeping fluency of the browsing, our summary browser combines several approaches described in the following paragraphs.

In the preprocessing phase, keyframes are detected in the endoscopic video based on tracking of keypoints with the KLT tracker [2]. We start with a dense grid of keypoints [6] and try to track them over time as long as possible. Due to motion in the video, more and more keypoints get lost over time and cannot be tracked further. As soon as the number of initial keypoints falls below a specific threshold (e.g., less than 50%), we extract a keyframe and restart keypoint tracking with a new grid of densely sampled keypoints.

Next, a hierarchical clustering is performed on the set of selected representatives resulting in a dendrogram that is used to assign a priority to each keyframe. More specifically, we employ an agglomerative hierarchical clustering where the cluster similarity is based on the Ward's method

$$\mathrm{L}_2^2(M(C_i), M(C_j)) \cdot (|C_i| \cdot |C_j|)/(|C_i| + |C_j|),$$

where C_i, C_j are two clusters, M(C) denotes the mean of all keyframes in the cluster and L_2^2 is the squared euclidean distance defined over two keyframes of the same size, where each pixel is represented by three values from the Lab color space. Furthermore, we allow just merging of two adjacent clusters, where the adjacency is based on the keyframe timestamp. For a new cluster, the timestamp is computed as weighted combination of two merged clusters. As a result of the clustering, a binary dendrogram for a set of representatives is obtained. Then the priority of each representative is assigned in such a way that the left most object (the lowest timestamp) in each cluster is assigned the highest priority.

The browser also employs timestamps for timeline-based visualization of actually visible top k representatives (based on their priority), where the actually

visible part/range of the video is controlled by a browsing user. The displayed images are pinned to the timeline based on their real position in the video. Let us note, placement of displayed representatives considers minimal overlap of the already depicted keyframes, because, as the representatives are distributed non-uniformly, a trivial placement technique would result in overlaps.

3 Dynamic Browsing of Clustered Keyframes

In order to browse the keyframes, the user can employ several browsing operations. As an initial view, the user sees the whole time-line period with top k representatives (Figure 1). At any position of this visualization, the user can use the mouse wheel for zooming in or out, which results in a smoothly animated visualization of more or less keyframes for the selected time region. After each zoom in/out operation, the middle timeline panel adjusts its border timestamps while the top and bottom panels just depict the temporal position of the actually visible range using black rectangles. In order to navigate the user to parts with yet invisible images, the timeline panels depict also positions of the images using gray circles.

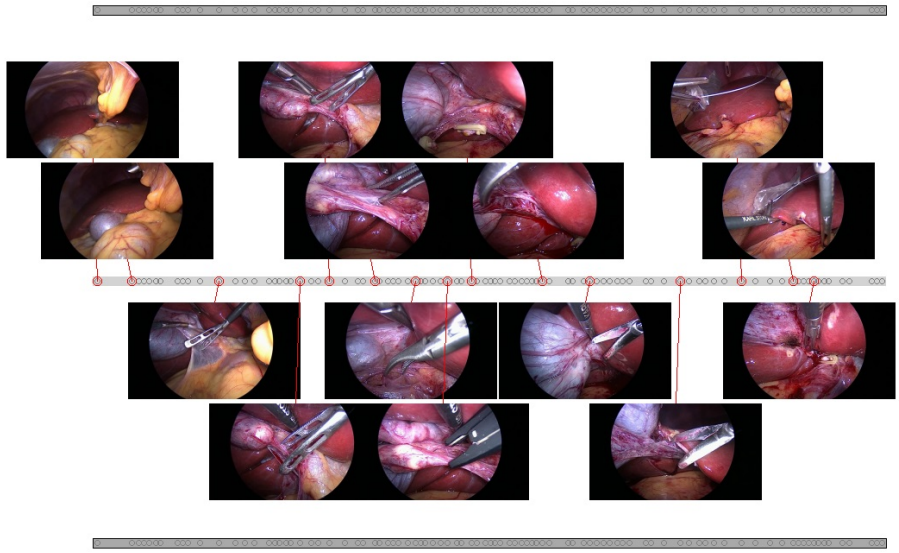

Fig. 1. Browsing endoscopic videos through an interactive hierarchical visualization of keyframes.

The zoom out operation is implemented as simple undo operation. The zoom in/out operations try to preserve the context of the browsing using animations that move already displayed objects to their new locations (some may disappear

if they reach the border) and after the objects are moved, the new objects with a high ranking in an actually visible range are placed and displayed. The user can also simply shift left/right using simple left/right mouse click, shifting the border timestamps by a given constant.

4 Conclusions

In this demo paper we have presented a video browser for endoscopic videos, which is based on a novel extraction and visualization of keyframes. The keyframe extraction uses significant motion changes in order to detect distinctive frames in the whole endoscopic video containing highly similar content. The visualization uses a compact layout of keyframes that shows the temporal positions the currently displayed keyframes as well as an indication of further available keyframes at higher levels of detail. A user can navigate within the tree of keyframes by using the mouse wheel and thus navigate deeper into the hierarchy for a specific temporal segment. The proposed video browser conveniently allows (1) to see a quick overview of representative frames in endoscopic video and (2) to see the details for specific temporal regions when selected by the user.

Acknowledgement

This work was supported by Lakeside Labs GmbH, Klagenfurt, Austria, by Czech Science Foundation project P202/12/P297 and by funding from the European Regional Development Fund and the Carinthian Economic Promotion Fund (KWF) under grant KWF - 20214 22573 33955. We would also like to thank to AKTION Czech Republic – Austria programme.

References

1. W. Hürst and K. Meier. Interfaces for timeline-based mobile video browsing. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 469–478. ACM, 2008.
2. B. D. Lucas, T. Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.
3. B. Münzer, K. Schoeffmann, L. Böszörmenyi, J. Smulders, and J. Jakimowicz. Investigation of the impact of compression on the perceptional quality of laparoscopic videos. In *Proceedings of the 27th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2014)*, pages 1–6, 2014.
4. J. Peng and Q. Xiao-Lin. Keyframe-based video summary using visual attention clues. *IEEE MultiMedia*, 17(2):0064–73, 2010.
5. K. Schoeffmann, M. Del Fabro, T. Szkaliczki, L. Böszörmenyi, and J. Keckstein. Keyframe extraction in endoscopic video. *Multimedia Tools and Applications*, pages 1–20, 2014. to appear.
6. J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 593–600, Jun 1994.