Panoramic e-Learning Videos for non-Linear Navigation

Rosália G. Schneider∗
Instituto de Informática
UFRGS

Manuel M. Oliveira†
Instituto de Informática
UFRGS

Abstract

We introduce a new interface for augmenting e-learning videos with panoramic frames and content-based navigation. Our interface gradually builds a panoramic video, and allows users to navigate through such video by directly clicking on its contents, as opposed to using a conventional time slider. We demonstrate the effectiveness of our approach by successfully applying it to three representative styles of e-learning videos: Khan Academy, Coursera, and conventional lecture recorded with a camera. The techniques described provide more efficient ways for exploring the benefits of e-learning videos. As such, they have the potential to impact education by providing more customizable learning experiences for millions of e-learners around the world.

CR Categories: I.3.6 [Computer Graphics]: Interaction techniques—;

Keywords: e-learning, interaction techniques

1 Introduction

There is a recent trend towards the online publication of educational material from universities around the world. Websites like Coursera [Cou 2012], edX [edX 2012], MIT OCW [OCW 2002] and Khan Academy [Kha 2009] are only some examples of a huge movement to make world-class education available to everyone, for free. The potential impact is huge: top universities are providing high-quality educational content to millions of students worldwide. In this work, we improve interaction with videos to be more customizable and effective. First, we create a panoramic representation of the video. Suppose the camera focuses on a given part of the board. There is no reason why the user should not be able to see the rest of the room, if this information can somehow be extracted. The availability of this panorama also allows us to navigate to explanations of the displayed content. For this, we use hyperlinks connecting the contents the user sees on the board to the exact points in the video where they were introduced. To the best of our knowledge, our work is the first to propose this navigation style. Additionally, our interface allows the user to annotate the video and export the panorama as a set of lecture notes.

One of the main challenges is the variety of styles of e-learning videos available on the web. We explored three types of videos, that we believe cover the majority of lectures available on the Internet: Khan Academy’s tablet videos, Coursera’s slide-based videos, and regular video lectures. These three categories are very different from each other and the applicability of our technique to them demonstrates the generality of the concepts.

The contributions of our work include:

• An interface to provide better interaction with video lectures that is applicable to existing videos;

• The first technique for content-based navigation through elearning videos.

2 Related Work

One of the first works to investigate interactivity in e-learning videos was developed by [Zhang et al. 2006]. They discuss a system based on a paradigm called learning by asking (LBA). The idea is to have short videos answering one specific question at a time. The authors use a short video explaining each slide of a lecture. The user is able to navigate by slide. One of the main restrictions is that it is necessary to generate and segment the videos in advance - we cannot apply this technique to existing videos.

A more recent work in interaction with video lectures is DragonFly [Corsten 2010]. DragonFly is a system to spatially navigate
through video lectures. The videos must be recorded using proprietary software to allow spatial navigation later. Also, the way they navigate in space is not by understanding the spatial relations among objects - but by artificially creating these relations. Finally, the testers for DragonFly explicitly asked for direct manipulation on the video [Corsten 2010].

Of course, our work is also related to research concerned with panorama creation and non-linear navigation of videos. Since our contribution is not in this area, we will not review this material, and will only refer to the specific tools and techniques we used in our implementation.

3 Online Learning Systems

Just like regular courses, online courses come in very different shapes. We divide our discussion into three main categories, according to the kind of video provided: canvas-based videos, slide-based videos, and recorded lectures. Most existing e-learning systems can be mapped into one of these three categories.

Canvas-based videos: In this kind of video the elements are typically text, equations, and drawing made with some sort of stylus. The background is usually of a fixed color, and the information is drawn with high contrast relative to it.

Slide-based videos: The background of the video is composed by a set of slides. On top of it, the instructor makes notes or marks specially important parts. Occasionally, the instructor is also shown in a small portion of the slide.

Recorded conventional lectures: One of the most common ways to make lectures available online is by simply uploading recorded lectures. While that is not the mostly adopted type of video for new courses, many universities have done that in the past. The variability is huge, making the automatic handling of all such videos a difficult problem.

4 Interface

The key contribution of this work is the idea of a user-friendly interface for providing more effective navigation and interaction for e-learning videos. We discuss the features it provides to the user.

Panoramic video:

We introduce the extended classroom metaphor. A student sitting in the classroom would be able to see things not shown by the limited field of view of the current video frame (red rectangle shown in Figure 2). Thus, we show e-learners panoramic video frames depicting the current content of the board/canvas, and providing immediate access to all content previously shown during the lecture (including erased content). Figure 2 shows one example of panoramic video frame produced by our system.

Direct access to content: A key concept in our design is the notion of non-linear navigation through a pointing interface. The user directly navigates to point where some concept has been presented by just pointing and clicking on it, on the panoramic video.

Automatic generation of notes: The panoramic video shows the whole content of the lecture, which can be exported as a set of lecture notes. Our interface also allows the user to annotate the video at any time, and export these annotations together with the notes.

5 Creating Panoramic Videos

To create a panoramic view, we need to find a projective transformation from the individual video frames to a global coordinate system. We will discuss the approaches for canvas-based videos, slide-based videos, and recorded lectures separately. In all cases, for the feature detection, description and matching, we used a GPU implementation of SIFT [Wu 2007] [Lowe 1999]. Feature detection is performed separately to each frame of the video.

5.1 Canvas-based videos

Canvas-based videos are the easiest to adapt to our technique. They have intrinsic spatial relations, high contrast and no occlusions.

Matching: In this type of video, there is a clear spatial relation from one frame to the next. Subsequent frames have significant overlap when transformed to a world coordinate system. Having significant overlapping regions allows for robust registration. In this case, we match the features in frame $i$ to the ones in frame $i-1$.

Homography: Once the features have been properly detected and matched, we need to estimate a homography between the sets of points from pairs of subsequent frames. We estimate the homography using DLT [Hartley and Zisserman 2004]. We use RANSAC to ensure the estimation is robust to outliers. We estimate a homography for each group of four pairs using the SVD implementation from CHOLMOD [Chen et al. 2008]. To create a panoramic frame, we multiply all homographies together and render the transformed frame over the existing ones. This is

5.2 Slide-based videos

Slide-based videos do not have explicit spatial information. Again, they do not suffer from occlusion problems.

Matching: When dealing with slide-based videos, the approach discussed in Section 5.1 is not effective since slides do not overlap. However, this type of lecture consists of a set of slides that can usually be downloaded with the video. In our system, we import these slides and place them side-by-side to create a big panoramic map of all content in the video. Once this is done, we match the features of each frame to this map.

Homography: In this case, we use a simpler approach to define the projective transformation of each frame. From our previous step, we already have the noise-free information of the position for the slide. Instead of estimating a homography, we calculate which slide has the most matches to the current frame. This strategy results in a fast and robust solution.

5.3 Recorded Lectures

Recorded lectures are the most challenging ones. One problem results from the fact that the boards are often occluded by the instructor. Our results show, however, that it is possible to create panoro-
mas and even to perform navigation in some cases, without taking special care of these issues. We treat a recorded lecture simply as a canvas-based video and use exactly the same algorithm described in Section 5.1.

5.4 Erased Boards

In canvas-based videos and recorded lectures, content will sometimes be erased to give place to newer formulations. When this happens, we want the old content to be kept accessible somewhere. For this, we allow the user to move new content (soon to be rendered), thus extending the panoramic video frame. We do not detect automatically when a board is erased. We have tested several simple metrics to decide whether a new frame shows an erased board, such as the number of features, but they are not robust enough to handle all cases in a satisfactory way.

6 Creation of hyperlinks

The ability to directly go to any point in the video with the use of hyperlinks is the most important extension provided by our technique. To create the hyperlinks, we rely on data already used for registration.

For each location in our reference coordinate system, we store the moment when a feature (e.g., some text or an equation) first appeared, and use it to navigate back when that content is selected. To avoid excessive cluttering, we segment our clickable map in 10x10 regions with at most one feature each. The stored features for a video segment can be seen in Figure 3.

Figure 3: Creation of hyperlinks. The red dots represent feature points with associated hyperlinks stored at those particular locations. Note how there is at most one feature per cell.

One issue that comes from non-linear navigation is that when we navigate back in time, for example, to the beginning of the video, we lose part of the content that was visible in the panorama (Figure 4(b)). While this is the expected behavior, it causes a problem when the user wants to return to the point where he/she was, or even to other explanations in between. To solve that, we give the user the option of seeing all the content shown so far by pressing a key. This "directory" is shown in gray shades (Figure 4(c)) to differ from the actual video frames. By seeing the hyperlinks position, the user can navigate back to the desired moment in the video.

Another issue with allowing non-linear navigation is that we do not store the frames of our panoramic videos - instead, they are created by rendering subsequent frames on top of each other. If we jump in our navigation, this will not create a panorama (Figure 5(a)). Also, when we navigate back in time, we do not want to keep future content visible (Figure 5(b)). To prevent that, every time our registration shows that a frame has moved significantly, we keep a panorama base for that moment saved in a buffer. The panoramic frame for any moment in the video can be restored using the panoramic base + the regular frame.

6.1 Automatic Generation of Notes

Our system automatically generates lecture notes from the panoramic video frames. For this, we simply export the panoramic view of the lecture together with the annotations the user made during the lecture. To handle PDF files robustly, we used the Poppler library, that integrates nicely with the Qt Framework we are using for the interface.

7 Results

This Section illustrates the use of our technique applied to the three classes of e-lectures. Figure 6, shows a panoramic frame and some of the regular frames used to create it. Note how the final panorama does not have artifacts, even in the presence of significant zooming (Figure 6, last frame).

Figure 6: Some frames (left) used to create the panorama (right). Note how the system naturally handles camera zooming.

Figure 7 shows the results obtained when using our technique with
a canvas-based style video. The image on the left shows the reconstructed panorama, highlighting three hyperlinks (a), (b), and (c) (shown in yellow on the left image). The images on the right show the frames associated with the corresponding links. The position of the cursor (highlighted in blue) in these target frames shows how our interface precisely navigates to the moment the content was drawn.

Figure 7: Results for a lecture on probability from Khan Academy. (left) Panorama with three hyperlinks highlighted in yellow. (right) Corresponding target video frames.

Figure 8 shows content-based navigation in a conventional lecture recorded with a camera. One can see that in each of the target video frames the instructor is about to write the content that was selected.

Figure 8: Results for the Synapses, Neurons and Brains Lecture on Coursera. (left) Panorama with highlighted hyperlinks. (right) Target video frames - note that the instructor is about to write the selected content.

7.1 Limitations

Our current prototype has several limitations. Having a robust registration step is essential for the creation of visually pleasing panoramas and to allow our hyperlinks to navigate to the correct moment in the video. The movement of boards was not treated in our implementation, and videos with labels result in problems in the panorama, when the label is rendered over the existing frames. These issues, however, are related to our current implementation and are not inherent to the concept we are introducing.

8 Conclusion

We presented a new interface for augmenting existing e-learning videos with panoramic frames and content-based navigation. Our technique also enhances the notion of a video player by allowing non-linear navigation through a simple pointing interface. We have demonstrated the effectiveness of the described concepts by applying them to three distinct classes of e-learning videos. These concepts provide more efficient ways for exploring the benefits of e-learning videos. As such, they have the potential to impact the learning experience for millions of e-learners around the world.

Acknowledgements

This work was sponsored by CAPES and CNPq (grants 482271/2012-4 and 308936/2010-8).

References


