A Two Tier Approach to Chalkboard Video Lecture Summary

Solomon E. Garber, Aaditya Prakash, Nick Moran, Maria Altebarmakian
Richard Alterman, Antonella Di Lillo and James A. Storer
Department of Computer Science
Brandeis University
Waltham, Massachussets 02453
Email: solomongarber@brandeis.edu

***Abstract*—We present a new tool to help students make use of video recordings of traditional chalkboard and whiteboard lectures. Our tool provides a compact summary of a chalkboard lecture in the form of a set of slides, and an interactive environment for electronic note-taking in real time while watching a lecture in class or during review.**

Key to the processing of a chalkboard video by our system is the removal of the lecturer from the video (now no longer present to occlude the board) and segmentation of the resulting video to produce a set of slides comprising a compact presentation of the content written on the board during the lecture.

Our method enhances learning in a number of ways. Students can focus on taking notes to supplement the material, rather than rote transcription. In addition, notes are aligned temporally with the lecture video and thus provide a useful way for students to index into the important parts of the video. Our system can be easily integrated into courses with a recorded lecture video component to help students organize their notes and quickly navigate the lectures from a given course for particular topics.

I. Introduction

With the increasing role of educational tools and resources made available by educators and institutions on the Internet, video recordings of classroom lectures are gaining prominence. Educators often upload recorded videos of their lectures to the Internet as an aid for their students, a part of the required coursework, or, in the case of the flipped classroom, as a replacement for the classroom lecture component of the course. Moreover, many lecturers publish these videos to a broad Internet audience. For these reasons, a need has developed for software that can help students navigate large volumes of lecture recordings and integrate the note-taking and review process into an electronic environment where videos can be viewed. Allowing students to have access to video of lectures has tremendous value. Lecture videos give students control over the rate at which the material is covered, and constitute a useful study resource.

Despite their value to students, series of lecture videos can be somewhat difficult to navigate, and do not serve as a substitute for notes and slides provided by the lecturer or notes taken by the student. Without some kind of structure, it is difficult to access a particular point in the lecture. Video titles and descriptions can not capture the breadth of topics covered in a typical lecture, and do not contain any of the key visual content.

The aim of this work is to make lecture videos more easily searched, reviewed, and annotated by students, without constraining lecturers in their teaching methods. In particular, this paper targets lecture videos where figures are drawn by hand onto a chalkboard (or whiteboard).

Students have limited options for navigating these videos. Video scrubbing (cursor-driven visual search) is a slow and faulty process for locating information in such videos. A simple solution to this problem which could apply to PowerPoint lecture video recordings would be to segment the video using the lecturer's slides: in other words, the student picks a slide they want to review and the system automatically moves to the portion of the recording where that slide was being projected. Synchronizing or extracting slides from video is a challenging problem in itself [1] [2]. Here we address the additional complexity of synchronization and extraction of slides from a chalkboard lecture video where the lecturer is blocking different portions of the board at different times, and the transition points are not as clear or sudden.

Popular approaches to slide extraction and segmentation for slideshow presentation videos can not be directly applied in the case where the lecturer makes use of visual aids such as a chalkboard, as is often the case for STEM courses. Firstly, when the lecturer gives chalkboard lectures there are no premade electronic slides that can be used to segment the video. Secondly, the majority of the visual content in these videos is the result of the activity of the lecturer in the foreground, and has little to do with the content of the lecture. Additionally, the board is often occluded by the lecturer, whereas this is rarely the case for slide based presentations.

The techniques presented in this paper make it easier to navigate a video of a lecture by segmenting the video. For a chalkboard lecture, the methods presented in this paper enable the system to remove occluding objects from the learners view of the chalkboard on the recording of the lecture, and to segment the lecture into slides.

The system presented here takes the burdensome part of note-taking off of the student (rote transcription of the contents of the board, or synchronizing slides and notes to video in the case of slideshow presentations), leaving room for more active note-taking and mental engagement with the material without requiring extra effort on the part of educators. As a

978-1-5090-5920-1/17/$31.00 ©2017 IEEE

further benefit, this system can be used retroactively on videos recorded without our process in mind.

In this paper we present a two tier representation of chalkboard-based lecture videos. We have designed a system that takes an ordinary chalkboard lecture video and creates two outputs, a slideshow summarizing the content from the lectures, and a video notepad - a shortened video where the lecturer has been removed - upon which students can write notes. The algorithm for extracting the video notepad from the input video can be run in real time, allowing students to take notes while the lecture is being given. We further process this shortened and unoccluded (with the lecturer removed) video notepad to obtain a minimal set of still images (or "slides") captured before anything is erased from the board. This slide set provides a compact summary of the contents of each video which can be used to quickly search many videos for a particular slide, or can even be used as a study guide or reference as one would use lecture notes. This hierarchical representation facilitates fast navigation of large volumes of video lectures, efficient note-taking promoting generative learning, and can also be used as a reference for the written contents of the videos, akin to a slideshow. We also present empirical results validating the proposed segmentation algorithm.

Section II contains a review of the literature on note-taking, on the strengths and weaknesses of chalkboard- and PowerPoint presentation-based lectures, and on the advantages and deficits of e-lectures and recorded lecture resources. Section II also contains a description of other systems for electronic note-taking, and automatic lecture summary and indexing. Section III describes how the lecturer is removed from the video feed to produce the video notebook. Also described in Section III is a new lecture video segmentation algorithm for generating slides and segmenting the lecture from the video, and in Section IV, the results of this segmentation algorithm are evaluated quantitatively. In Section V, the functionality of our system is demonstrated with a proposed user interface and workflow, and future directions for this work are discussed in Section VI.

II. Background and Motivation

Instruction via in-class lectures remains the default method of teaching in universities. [3] estimated that 80% of class time at the university level was spent in lectures. Still, there is great variation in teaching techniques in terms of lecture style, course structure, and the quantity and quality of materials made available to students outside of the lecture. Many studies, discussed below, have examined the possible impacts on student perception and academic performance of different lecture styles, course structures, as well as note-taking and studying practices on the part of students.

a) The flipped classroom: Flipped classrooms, where students view lectures outside of class as a recording and attend class to get help on their work, have been discussed extensively. In [4], a controlled study found that students in a Human-Computer interaction course got better grades in the flipped classroom setting than students in the traditional in-class lecture group. [5] found consistent results with test scores, and additionally noted that students provided with a lecture video to watch outside of class came to class better prepared than those in the control group. Our technique lends itself to the flipped classroom format by making it easier for professors to curate and students to use lecture videos. Our methods can also apply to any course where lecture videos with a chalkboard or whiteboard component are made available to students. [6] shows how the flipped classroom format can be beneficial for engineering students developing lifelong learning skills. They emphasize student control of learning and pacing as helpful for promoting lifelong learning. By helping students navigate volumes of digital lectures and organize their own notes, our system facilitates self-driven learning at the user's pace. Our technique could further improve lifelong learning by allowing students to design their own curricula by searching through lecture videos from various courses, focusing on the topics they deem most important, as explained in Section VI.

Many educators, particularly those in STEM fields, rely heavily on visual aids in their lectures. Generally the verbal component of the lecture is accompanied by either black/whiteboard, overhead projector, or PowerPoint slide presentations. These various methods have been competing for decades, and many studies have been done to document the strengths and weaknesses of each method [7] [8] [9] [10] [11] [12] [13] [14].

b) PowerPoint: In a study conducted by [7], it was found that students of Environmental Science preferred PowerPoint lectures to overhead projector based lecture, and those students also scored better on their exams. [8] found results consistent with those of [7] using two different experimental setups, although they ran a third experiment which somewhat undermined their claims. [9] verified students' opinion of PowerPoint using a questionnaire; students reported that PowerPoint lectures helped them learn and focus on the presentation.

The PowerPoint lecture method has also faced some criticism, and not all experimental results have given such a positive picture. [10] argues that PowerPoint lectures prevent meaningful interactions between the lecturer and the audience.

[11] criticizes PowerPoint presentations from multiple angles. Firstly, that the process of creating concise, attention-grabbing slides emphasizes slogans over detailed explanations. Secondly, that the flexibility of the software allows users to include overcomplicated and confusing figures and graphics, and make formatting choices which are distracting to the audience. [11] also claims that the linear step-by-step nature of PowerPoint presentations is a hindrance when presenting complicated topics with many connections.

[12] and [13] both cite evidence that lecturers have a more monotonous tone when reading from a slide than when speaking directly to an audience, and they give fewer nonverbal cues to their audience to help filter the most relevant information. A study done by [14] supported these ideas. They found that students scored lower on tests where the lectures were given

by PowerPoint presentations, despite confirming [7] and [9]'s finding that the students prefer PowerPoint over traditional lectures.

[13] found a very strong correlation between PowerPoint lectures and decreased test performance. They suggest that the reason for the discrepancies in the results of different authors are a result of differing experimental setup. Authors who found a positive impact of PowerPoint presentations were using it not as the central focus of the lecture, but as an auxiliary tool to improve their traditional lecture. On the other hand, the authors who found negative impacts were testing the efficacy of PowerPoint-centric lectures with very little deviation from the content of the slides.

Our technique allows teachers the flexibility to give extemporaneous chalkboard lectures while capturing a useful slide summary for supplementary use by students. While the question of the best visual media for student learning is still an actively researched area, our technique seeks to provide instructors with flexibility in their selection of visual aids.

c) Note-Taking: A parallel and related discussion has been happening over the importance of note-taking and the use of digital notes provided by the lecturer. It is no surprise, as shown by [15], that students who take better notes perform better on the end-of-class quizzes. [15] argues that note-taking is an example of active engagement of students with the material, and in some cases, when students include their own thoughts and connections in their notes, an example of generative learning. [3] points out that most students take highly incomplete notes, and gives evidence that more complete note-taking is correlated with better performance.

By providing digital notes and slides, lecturers can free students of some of the burden of copying and allow them to engage in more generative note-taking practices. [16] tested the efficacy of lecturer provided notes. In their study, some participants were provided with complete lecture notes, some with a note sheet containing just the major headings of topics discussed, and the third group was not provided with any notes. Both groups that received notes outperformed the no note group, and interestingly the group with the more limited notes did the best at tasks involving transferring knowledge. Both groups with notes did equally well at remembering facts.

[17] highlights more results supporting the claim that students who are given skeletal notes to annotate during lectures outperform students who are given more detailed notes and students who are left to their own devices. The authors suggest that skeletal notes allow students to take more complete notes and encourage students to draw connections on their own while taking notes. Our system relieves students of having to copy the lecturer's writing exactly, allowing them to write more of their own notes and connect them directly to the natural skeletal notes generated by a lecturer writing on the chalk board.

d) E-Learning and the recorded lecture: In addition to electronic notes, access to technology in the classroom has allowed lecturers to easily publish videos of their lectures for their students to review. Many lecturers, as well as educational institutions including Stanford, UC Berkeley, IIT, and many more, have made their lecture videos available to the broad public. [18] points out that video lectures and e-learning environments allow students to set their own learning pace and can give the audience a more active role in the learning process. [19] found that students were more likely to review recorded lectures than to read supplementary materials, and that they were more likely to take notes while reviewing recorded lectures than assigned readings.

Despite their clear value, online volumes of lecture videos pose many challenges to students attempting to use them. The main challenges in creating online environments for video lecture e-learning, according [18], are in facilitating fast browsing and locating content of interest, linking related content (like the slides associated with a particular lecture), and allowing users to collaborate with one another. [20] points to efficient browsing and searching interfaces and improved content based video indexing as important research problems for video centric e-learning. We generate the supplementary slides automatically from the video, so they can easily be used as an index into the video. Our video notepad also gives students a convenient place to record notes, and these notes can be easily searched and can act as a temporal index into the lecture.

e) Improving Navigability: Many different approaches have been taken to address the problem of navigability of video lectures. Projects like the Intelligent Classroom [21] combine hardware and software solutions to facilitate lecture recording and publishing among teachers and to improve the quality of these videos. Along the same lines, [22] presents an end-to-end system which allows lecturers to record and publish videos of their lectures without the need of technical assistance. [23] proposes a method for automatically controlling a pan-tilt-zoom camera, removing the need for a camera operator even for large lecture halls that do not readily fit into a camera frame.

Another strategy for making video lectures easier to navigate is to segment the videos based on some measure of change. Generally, the hope is to automatically separate the video into meaningfully distinct segments. [24] uses optical character recognition (OCR) to extract text from a slide presentation, voice recognition to create a transcript, and text-based analysis techniques to detect topic changes. [25] presents another text-based method for segmenting lectures by topic, exclusively using speech from the video. By contrast, [26] generates video segments directly from the video feed, in their case, video of slide presentations. Our work is more in keeping with the video based segmentation approach of [26], but in our case the input videos are of chalkboard and whiteboard lectures, posing significant additional challenges.

In many instances, there is a set of PowerPoint slides accompanying a lecture video. In this situation, segmentation can be achieved simply by linking the slides to the part of the video where they are shown. This is the approach of [1], in which OCR techniques are used to locate the slides in the video, and [2], in which low level image features are used.

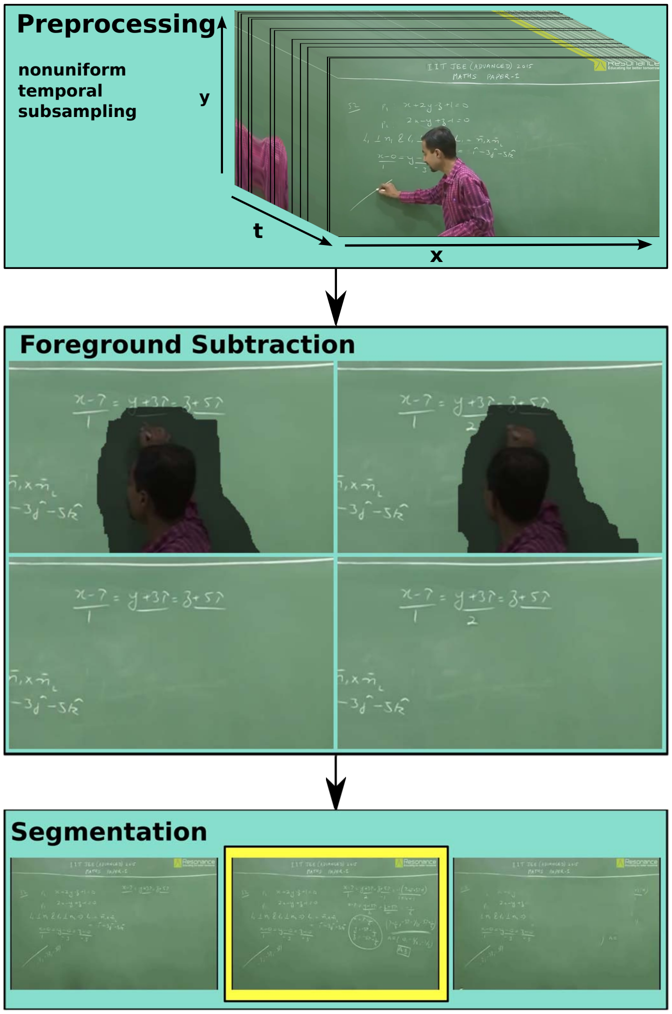

Fig. 1. Diagram of the video processing pipeline for removal of the lecturer from the video. In the preprocessing stage, a minority of frames from the original feed are selected for further processing. In the foreground subtraction stage, foreground masks are used to block out the lecturer in regions of motion. In the slide selection block at the bottom, the middle frame has been selected as a slide because the next frame has some information erased. Best viewed in color.

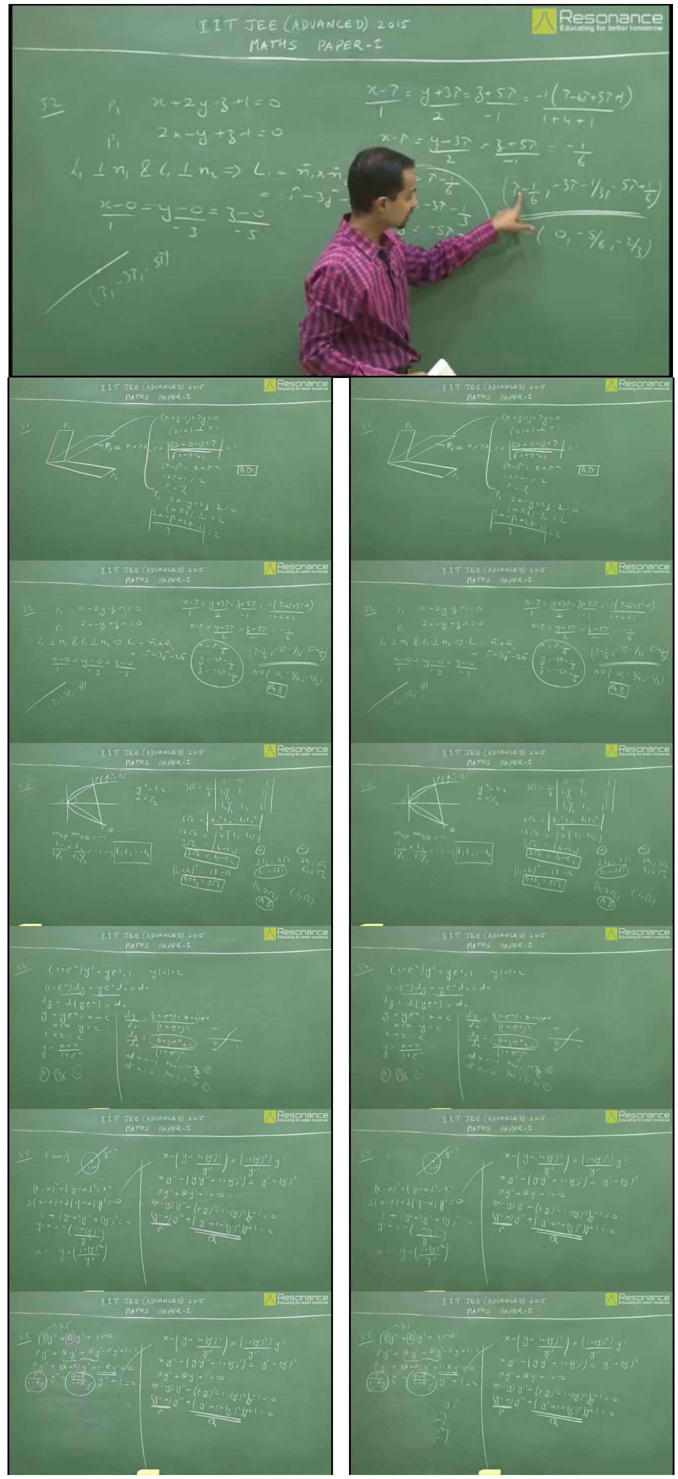

Fig. 2. A sample input frame(top), ground truth slide set(left) and corresponding slides produced by the system described in this paper(right). Another unnecessary slide was selected for this video, shown in Fig. 4. Best viewed in color.

Our work automatically segments videos by their slides.

Text extracted from video and audio of lecture presentations can also be queried by students trying to locate a particular topic and used as an index into the video. In [26] and [27], text is extracted from the slides in the video for this purpose. [28] takes this a step further by generating a text transcript as well as extracting text from the slides in the video. In [29], handwritten digits are extracted from mathematical lecture videos by combining the audio recognition with handwriting recognition. Our approach allows students to use their own notes as an index into the video.

Some authors have generated slides or other visual summaries directly from video of chalkboard or whiteboard presentations, much like our system. In [30], a method is presented for stitching together photos of a whiteboard presentation, removing the speaker and extracting the the pen strokes for low bandwidth transmission of whiteboards in remote collaboration settings. [31] uses a similar technique to classify chalk pixels in chalkboard lecture videos. These classification results are then used as a black and white slide, and the video is segmented at times when there is a local maximum in the number of pixels classified as chalk. [32] uses video processing techniques to extract writing from a whiteboard presentation. Segmentation is performed by a method similar to that of [31], with post processing techniques applied to the slides to sharpen the written characters. OCR is then applied to the result.

Some authors have attempted to extract notes from blackboard-style videos (instructional videos recorded by *screen capture* in a drawing application, as opposed to a natural video). In the work of [33], transcripts are extracted from the audio feed, which are then interleaved with visual elements from the video to make a visual transcript which could theoretically be used in lieu of watching the original video, and certainly for searching, indexing and review. In a similar vein, [34] turns blackboard-style videos into an interactive environment where each written element from the video is visible, and can be clicked to play the relevant portion of the audio where the character was written. Neither of these works targets natural video, and none of the aforementioned works allow students to add their notes directly to the videos.

In the LiveNotes system [35], an environment is presented where students can annotate PowerPoint presentation slides with notes and connections while the slideshow is being presented. Chalkboard lectures have many advantages over PowerPoint presentations in terms of flexibility for the lecturer and interaction with the audience, but video recordings of chalkboard lectures (the target of this work) pose many challenges to students trying to keep their notes organized or review the material presented in a course. Our two tier system both allows users to annotate the chalkboard during the lecture, and automatically generates slides for later studying and review.

III. Our Method

We process the video in multiple stages to produce our video notepad and minimal slide set (see Figure 1).

A. Preprocessing

First, because lecture videos contain a lot of visually redundant information, we create a "time lapse" version of the original video by only considering a subset of the frames from the video. The amount of time we let elapse between sampled frames is variable and depends on the amount of change in the pixel values of the video – see [36] for the technical details of this method. This helps make a more compact summary for scanning purposes, and allows our algorithm to generate unoccluded frames in real time.

B. Foreground Subtraction

We further process the time lapse video to remove sudden changes, which are generally due to the lecturer or a student occluding information on the board. We extract the background from the time lapse video by masking out the regions of change between frames, and only updating pixel values in the stable regions. We use the Pearson correlation coefficient between pixel values in temporally adjacent blocks to determine when a change has occurred. Because Pearson correlation can be written in terms of expected values, we can compute it efficiently over the whole frame using uniform blur kernels, which are separable (i.e., the 2D kernels can be written as the outer product of two 1D vectors, and filtering can be implemented as two successive convolutions by 1D filters rather than a single convolution by a 2D filter). This means the number of multiplies per pixel is on the order of the width of blur kernel, not the square of the width (as would be the case for a general 2D convolution). Since this computation depends only on the pixel values within a small neighborhood of each pixel, and because we have implemented our foreground subtraction using pointwise operations and separable linear filters, our computation time per frame can be greatly reduced with the use of architectures supporting SIMD instructions, such as a GPU. By keeping a background frame in memory, which we update as new frames come in from the video, we can generate lecturer-free slides on demand.

We blend the images at the mask boundaries to create a clean output video. This extracted background video contains all of the motionless regions from the original video, including the information written on the board (since this information does not change from frame to frame except when it is written or erased). For this reason, writing may not appear in the processed video immediately after being written, and its disappearance after being erased may also be slightly delayed. We found these effects to be negligible (on the order of a few seconds) in most cases.

The output of this first phase of processing is what we use for the frames of our video notepad. Because our foreground subtraction algorithm does not require a first pass over the video and can be run in real time, students can annotate frames from the video notepad while the lecture is being given. This video notepad is also the input for the second phase of processing, where a set of frames is chosen to be used as a slide set and video segmentation checkpoints. This segmentation algorithm requires multiple passes over the data, and so it can not be run in real time.

C. Segmentation

Although the peak finding approach used by [31] and [32] works well for segmenting, we found that it was often too sensitive to noise to pinpoint the best slide before the board is erased. Instead, we apply the similarity metric used in the first phase to erase the lecturer to detect when writing has been added to or erased from the board.

We take as slides any frame preceding a significant erasure event (past some low threshold). This approach reliably generates slides with all of the writing from the lecture, and it is able to take the last slide before the board is erased, which often means capturing key details that are added toward the end of a segment. It does have a tendency to generate false positives (extraneous slides) as a result of noise, lecturers taking a long time to erase, changes in lighting, and various other challenging environmental conditions. In addition, even working as designed, the system is prone to taking redundant

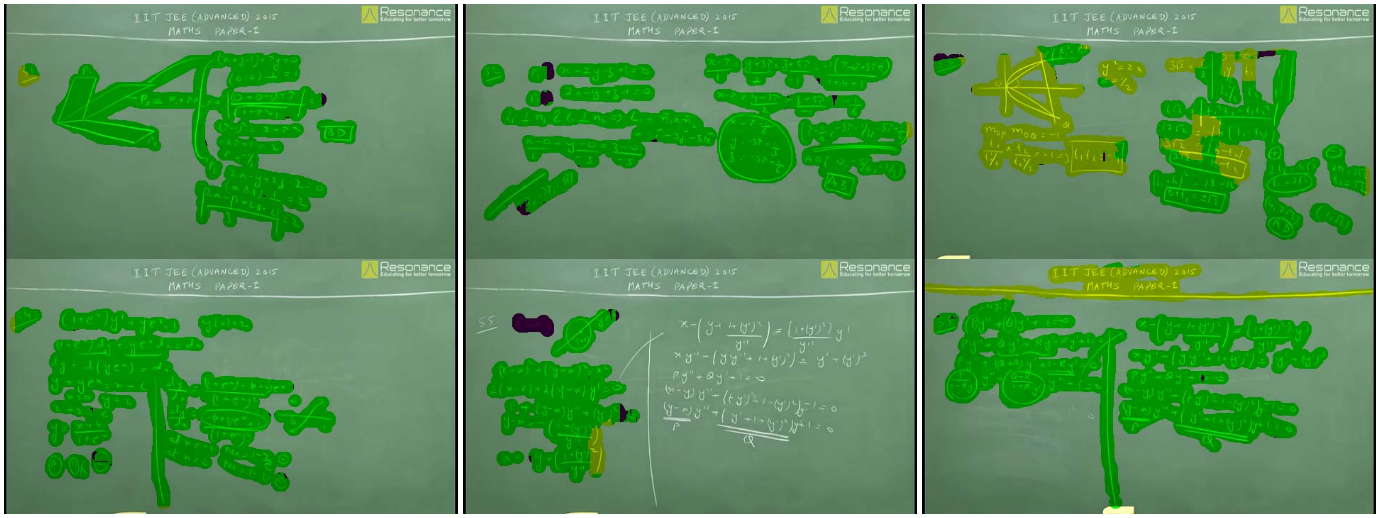

Fig. 3. Ground truth slides colored based on the redundancy in the segmentation output (Figs. 2 and 4). Green regions were in the generated slide set the same number of times as the ground truth slide set. Yellow regions were in the generate slide set once or twice more than the ground truth, and purple regions were omitted from the generated slide set. Because our algorithm only generated one extra slide for this video, there were no orange regions (more than two more than the ground truth).

slides as many lecturers tend to update and erase only half of the board at a time, leading to twice as many slides as necessary. However, this can be easily corrected. We process the slides one last time to remove the ones that are entirely redundant (each of their edges has a match in at least one other slide).

IV. Experimental Evaluation

Critical to the quality and usefulness of the system described is the completeness and parsimony of the slide set generated during segmentation of the unoccluded video. Since the slide set is important for skimming the contents of an unwatched video for a particular topic, and since they are meant to be usable as a review tool, it is important that they contain all of the information written on the board during the lecture. Also, for the segmentation results to be useful for separating chunks of video by topic, it is important not to create many extraneous slides in order to avoid potentially meaningless and distracting video segments.

A. *Experimental Design and Measurements*

In order to validate the proposed method, we collected 16 lecture videos “in the wild” (from YouTube) to be used as the basis for evaluation. For these videos to be useful as a means of testing the results of segmentation, we needed an optimal set of slides associated with each video - we will refer to this as the ground truth slide set. The ground truth slide set was created to be optimal in the sense that it contains all of the information written on the board during the lecture, and that no other set of slides containing all of the information could have fewer slides than the ground truth. For each video considered, the minimal number of slides necessary to summarize the whole of the information written on the chalkboard during the course of the lecture was determined. Slides were extracted from the video manually, blending images when necessary (using the blending technique of [37]). The left column of Fig. 2 is an example of a ground truth slide set, and the right column contains the corresponding slides generated by the proposed method. Fig. 4 shows the one redundant slide selected by the segmentation algorithm for the same video.

The number of images in this human-generated slide set is helpful for verifying that the segmentation algorithm is not generating spurious slides which could slow down studying, but it was also important for us to verify that the generated slides contained all of the information written on the board during the lecture.

Because the lecturer does not necessarily erase the entire board between each new slide, the ground truth slide sets often contain identical information on more than one slide. Furthermore, some objects visible in the slides may not be relevant to the content of the lecture (e.g. a chalkboard eraser or sticks of chalk), and these elements may be moved during the lecture, and thus appear in different locations in equally valid slide sets. In order to determine how much of the important information from the lecture was present in the slides (without over-counting regions which appear in multiple slides or penalizing placement of irrelevant objects), ground truth masks were constructed by marking the important regions of each ground truth slide. Anything written on the board is behind just one of the masks. Fig. 3 shows how these masks can be used to evaluate a proposed slide set, where each slide in the ground truth slide set has been colored according to the number of times it appeared in the automatically generated slideshow in comparison with the ground truth, where purple regions were not matched to any slide, green regions occur at least once but no more than in the ground truth, yellow regions one or two more times than in the ground truth, and red regions more than two times more than in the ground truth.

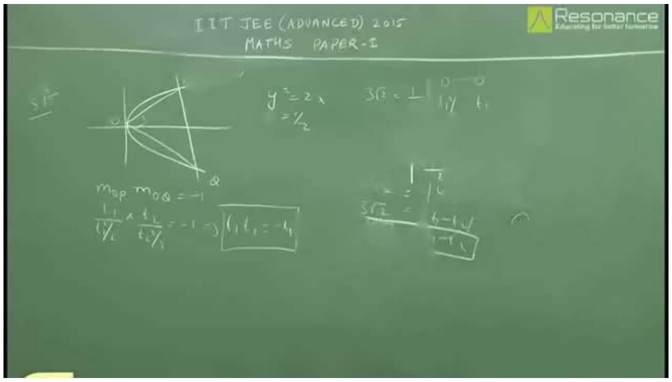

Fig. 4. The extraneous slide generated during segmentation for the slideshow in Fig. 2

B. Results

The two metrics we used to judge the slideshows generated by the tool developed here are $\mathcal{R}$ = the ratio of the size of a proposed slide set to the ground truth, and $\mathcal{P}$ = the percent of masked pixels with no match in the proposed slide set. For $\mathcal{P}$, smaller is always better because it means fewer pixels from the ground truth are missing. Across the videos in the ground truth, the mean value of $\mathcal{P}$ was 2.48%, with a standard deviation of 2.04, meaning in general, 95.48 to 99.56% of important pixels were not lost in the final sides. For $\mathcal{R}$, 1.0 represents the ideal. For $\mathcal{R} < 1.0$, some information must have been lost, for $\mathcal{R} > 1.0$, some slides must be extraneous. We did not have any videos where $\mathcal{R}$ took a value below 1.0, and across the test set we saw a mean $\mathcal{R}$ of 1.6 with a standard deviation of 0.67. The distribution of these values was somewhat right-skewed: we saw a median $\mathcal{R}$ of 1.33, so we generally produced about $1/3$ to $2/3$ too many slides. This is still somewhat high, and improving that score is an area of future work for us, but we intentionally prioritized completeness over brevity in slide sets when designing the algorithm so that students could use them as a study guide. The bottom line is that skimming through 16 slides instead of 10 is not that much extra work, especially considering that the alternative to searching an automatically generated slide set is scrubbing the video.

C. Evaluating the Results

The results of the experiment provide evidence that the method presented can reliably segment the instructional videos in a manner that generates a good approximation of a minimal and complete slide set. Without some kind of segmentation, the learner is reduced to video scrubbing. With the approximately minimal and complete slide set, the search time is reduced.

Many of the videos we tested were over an hour long with more than 10 slides in the human generated minimal ground truth slide set. If a user can see the particular equation they are looking for on a slide for such a video, they are relieved of scrubbing to find the relevant segment. At the worst, they will only have to scrub over a single segment - a small fraction of the entire video. It is often not clear to users from the titles or descriptions which video in a series contains a particular piece of information. The segments allow the learner to skim and see the overall structure of the lecture. This allows students to select which video is the right video, in addition to locating information within the video.

V. Discussion

As discussed above, taking notes during a lecture can be demanding of cognitive resources. The more work it takes to write the notes, the less mental energy that can be applied to thinking about the lecture (see [17], [3]).

With the scenario that is being developed in this paper, the student note-taking effort is greatly reduced during a chalkboard lecture. Suppose that a video feed of the lecture is available to each student as the lecture is given. Rather than write out lengthy notes, the student can periodically grab an unoccluded snapshot of the lecture and annotate. After class, students can use the video of the lecture as a basis for review. Because the notes were produced by taking "snapshots" of the video feed as it was produced, the notes themselves can be used to directly access the relevant parts of the video of the lecture.

Given a slide-based lecture, it would be possible for the lecturer to simply provide the slides directly to the students. However, this still leaves the problem of integrating the verbal content of the presentation and keeping the content unified for review purposes. The method presented in this paper can be used to identify the segment of video that corresponds to each slide during the lecture, allowing a student to jump directly into the corresponding section of the original lecture video. Other researchers such as [1] and [2] have addressed the problem of using slides to segment a video of a lecture. An advantage of the approach taken here is that it can segment slides and diagrams on the chalkboard simultaneously in instances where both are used, and it does not rely on OCR, so slides are not required to contain text.

The video notepad – i.e., the unoccluded video stream – allows users to listen and think about the lecture with confidence that users will be able to revisit the notes written on the board during the lecture at a later time.

This arrangement – providing the ability to rapidly access relevant material from the video – has a tremendous impact on note-taking during the lecture. Before it was feasible to record and distribute lecture videos, students were totally reliant on the quality of their own notes when reviewing a lecture. As discussed in Section II, there is a tradeoff between the time and effort devoted to taking notes and the ability to focus on absorbing the content. For this reason, enabling students to efficiently use the video as a backup greatly reduces the burden of note-taking. During the lecture, rather than attempt to take flawless notes, the student can mark a segment for review by taking a snapshot from the video notepad. Later, the student can access the part of the video stream relevant to this material. This setup gives the student access to the best of both worlds: she can take good notes and at the same time listen actively to, and think productively about, the lecture.

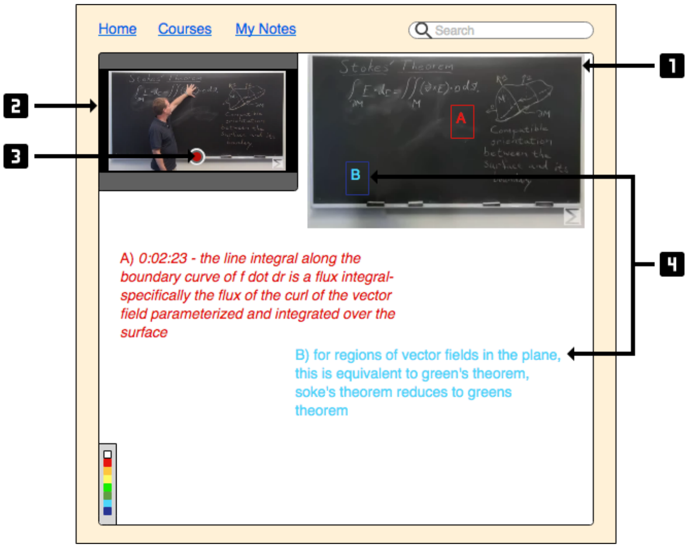

Fig. 5. A video notepad environment for taking notes during the lecture and reviewing the lecture recording and notes. 1) snapshot from the video notepad. 2) live stream or playback of the video lecture recording 3) record button for taking a snapshot of the unoccluded screen to start a new note. 4) Note tag for linking notes to the unoccluded snapshots, and the associated note.

The live interface for taking snapshots during a chalkboard lecture could be fairly simple. Suppose the students are virtually attending the lecture. The students would be looking at the video feed of the lecture and with push of a button could take a snapshot (see Figure 5). As shown in the figure, the user interface would display the video stream, the snapshot just taken and space to take notes.

[35] created a user interface for students to add their own notes to PowerPoint slides. The techniques described in this paper can give greater flexibility by allowing students to select when they want to create a slide.

The window in the left column of Figure 5 shows how such an interface, combined with the system described in this paper, integrates the lecture viewing and note-taking experiences in a digital setting. The upper left corner of this window shows an example frame from a live lecture video feed. The upper right corner shows a corresponding snapshot from the video notepad at the same time in the video. Two notes have been added to the snapshot. Since the foreground subtraction method described here can run in real time, this functionality can be made available to students during the lecture while the video is being recorded, or during subsequent review.

VI. Conclusion and Future Work

In this paper, we presented a tool to improve student engagement with video recordings of chalkboard and whiteboard lectures. Our tool can automatically generate compact slide summaries of lecture videos for review, and a helpful real-time video stream of the unoccluded chalkboard which can be used to take notes while watching a lecture in class or reviewing the lecture video afterwards. We showed that our method is not only able to remove the lecturer from a chalkboard lecture video, it can also segment the resulting video to generate small slide sets containing nearly all of the information from the board. In addition, we demonstrated that this can be done retroactively on videos recorded without our method in mind, meaning our system can be easily integrated into courses where lecture videos are already being recorded.

Our technique has many potential benefits to learning. Students can take fewer notes, allowing them pay closer attention to the lecture. Also, since students can easily use their notes as an index into the lecture video, they can focus their note-taking on recording their own thoughts and connections.

The methods discussed in this paper have been presented in terms of the traditional classroom lecture format. The impact of these methods on learning goes beyond the traditional classroom. For example, in a STEM-based workforce, workers often need to fill in gaps in their knowledge. They now have access to many lecture video resources from websites like YouTube, but review of the video can be time consuming. These methods can be used to speed up that process. Similar advantages accrue to lifelong learners, graduate and undergraduate students seeking supplementary educational materials on the web, medical and other students seeking professional degrees in technical fields, and others.

There are a number of areas for future research. The data we get from the slide selection component of our system could potentially be used to improve the segmentation algorithm or to generate sub-segment chapter proposals. Currently our segmentation algorithm only creates new segments when something is erased from the board. This is sufficient for generating a useful slide preview, but many lectures contain significant topic changes which do not coincide with information being erased from the board. The additional segments added by students help mitigate this problem. We intend to use the data from these user interactions to generate segmentation boundaries.

There are also several approaches by which we intend to analyze the impact of this work in the engineering classroom. We intend to measure the speed at which students can locate particular segments of a lecture with and without the aid of the video notepad, as well as assess the potential benefits to knowledge retention based on the test scores of classes instructed with and without the aid of the tools described here. We also intend to gauge student opinions of the tool directly through a questionaire.

In addition, we hope to extend our technique to make video notepads for more general lecture videos including pan-tilt-zoom, blackboard-style (Kahn academy style screen-cast) videos, and mixed media videos. Also, since publicly available lecture videos have become so popular, we seek to create a searchable database of lecture videos with video notepads, so that the interface can be used for self driven and lifelong learning as well as to supplement courses taken for credit (the target of this paper). This could allow students to design their own curricula based on their particular needs (by combining lectures from different content providers) and share their notes and favorite videos with classmates and others.

REFERENCES

[1] F. Wang, C.-W. Ngo, and T.-C. Pong, "Synchronization of lecture videos and electronic slides by video text analysis," in *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003, pp. 315–318.

[2] M. Eberts, A. Ulges, and U. Schwanecke, "Amigo-automatic indexing of lecture footage," in *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on*. IEEE, 2015, pp. 1206–1210.

[3] B. B. Armbruster, *Taking notes from lectures Book: Handbook of college reading and study strategy research*. Routledge, 2000, ch. 7, pp. 175–199.

[4] J. A. Day and J. D. Foley, "Evaluating a web lecture intervention in a human–computer interaction course," *IEEE Transactions on education*, vol. 49, no. 4, pp. 420–431, 2006.

[5] T. Stelzer, D. T. Brookes, G. Gladding, and J. P. Mestre, "Impact of multimedia learning modules on an introductory course on electricity and magnetism," *American Journal of Physics*, vol. 78, no. 7, pp. 755–759, 2010.

[6] L. Bland, "Applying flip/inverted classroom model in electrical engineering to establish life-long learning," in *ASEE Annual Conference & Exposition*, 2006.

[7] R. B. Lowry, "Electronic presentation of lectures-effect upon student performance," *University chemistry education*, vol. 3, no. 1, pp. 18–21, 1999.

[8] A. Szabo and N. Hastings, "Using it in the undergraduate classroom: should we replace the blackboard with PowerPoint?" *Computers & education*, vol. 35, no. 3, pp. 175–187, 2000.

[9] J. M. Apperson, E. L. Laws, and J. A. Scepansky, "The impact of presentation graphics on students? experience in the classroom," *Computers & Education*, vol. 47, no. 1, pp. 116–126, 2006.

[10] M. Martha Driessnack *et al.*, "A closer look at PowerPoint®," *Journal of Nursing Education*, vol. 44, no. 8, p. 347, 2005.

[11] E. R. Tufte, "Power corrupts. PowerPoint corrupts absolutely," *Wired*, p. 118, 2003.

[12] H. Maddox and R. Loughran, "Illustrating the lecture: Prepared diagrams vs. built-up diagrams," *Educational Technology Research and Development*, vol. 25, no. 1, pp. 87–90, 1977.

[13] R. C. Pros, A. C. Tarrida, M. d. M. B. Martin, and M. d. C. C. Amores, "Effects of the PowerPoint methodology on content learning," *Intangible Capital*, vol. 9, no. 1, pp. 184–198, 2013.

[14] N. Amare, "To slideware or not to slideware: Students' experiences with PowerPoint vs. lecture," *Journal of technical writing and communication*, vol. 36, no. 3, pp. 297–308, 2006.

[15] C. Stefanou, L. Hoffman, and N. Vielee, "Note-taking in the college classroom as evidence of generative learning," *Learning Environments Research*, vol. 11, no. 1, pp. 1–17, 2008.

[16] A. D. Katayama and S. M. Crooks, "Online notes: Differential effects of studying complete or partial graphically organized notes," *The Journal of experimental education*, vol. 71, no. 4, pp. 293–312, 2003.

[17] K. A. Kiewra, "A review of note-taking: The encoding-storage paradigm and beyond," *Educational Psychology Review*, vol. 1, no. 2, pp. 147–172, 1989.

[18] K. Hornbæk, D. T. Engberg, and J. Gomme, "Video lectures: HCI and e-learning challenges," in *Workshop on Human-computer interaction and E-learning*, 2002.

[19] G. Luna and D. Cullen, "Podcasting as complement to graduate teaching: Does it accommodate adult learning theories?." *International Journal of Teaching and Learning in Higher Education*, vol. 23, no. 1, pp. 40–47, 2011.

[20] T. Liu and J. R. Kender, "Lecture videos for e-learning: Current research and challenges," in *Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium on*. IEEE, 2004, pp. 574–578.

[21] L. R. Winer and J. Cooperstock, "The "intelligent classroom": Changing teaching and learning with an evolving technological environment," *Computers & Education*, vol. 38, no. 1, pp. 253–266, 2002.

[22] R. Kannan and F. Andres, "Towards automated lecture capture, navigation and delivery system for web lecture on demand," *International Journal of Innovation in Education*, vol. 1, no. 2, pp. 204–212, 2010.

[23] H.-C. Liao, M.-H. Pan, M.-C. Chang, K.-W. Lin *et al.*, "An automatic lecture recording system using pan-tilt-zoom camera to track lecturer and handwritten data," *International Journal of Applied Science and Engineering (IJASE) 13 (1)*, pp. 1–18, 2015.

[24] T. Tuna, M. Joshi, V. Varghese, R. Deshpande, J. Subhlok, and R. Verma, "Topic based segmentation of classroom videos," in *Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE*. IEEE, 2015, pp. 1–9.

[25] M. Lin, J. F. Nunamaker, M. Chau, and H. Chen, "Segmentation of lecture videos based on text: a method combining multiple linguistic features," in *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*. IEEE, 2004, pp. 9–pp.

[26] H. Yang, F. Gruenewald, and C. Meinel, "Automated extraction of lecture outlines from lecture videos-a hybrid solution for lecture video indexing." in *CSEDU (1)*, 2012, pp. 13–22.

[27] H. Yang, M. Siebert, P. Luhne, H. Sack, and C. Meinel, "Lecture video indexing and analysis using video ocr technology," in *Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on*. IEEE, 2011, pp. 54–61.

[28] H. Yang, C. Oehlke, and C. Meinel, "An automated analysis and indexing framework for lecture video portal," in *International Conference on Web-Based Learning*. Springer, 2012, pp. 285–294.

[29] S. Vemulapalli and M. Hayes, "Audio-video based character recognition for handwritten mathematical content in classroom videos," *Integrated Computer-Aided Engineering*, vol. 21, no. 3, pp. 219–234, 2014.

[30] L.-W. He and Z. Zhang, "Real-time whiteboard capture and processing using a video camera for remote collaboration," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 198–206, 2007.

[31] C. Choudary and T. Liu, "Summarization of visual content in instructional videos," *IEEE Transactions on Multimedia*, vol. 9, no. 7, pp. 1443–1455, 2007.

[32] L.-w. He, Z. Liu, and Z. Zhang, "Why take notes? Use the whiteboard capture system," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 5. IEEE, 2003, pp. V–776.

[33] H. V. Shin, F. Berthouzoz, W. Li, and F. Durand, "Visual transcripts: lecture notes from blackboard-style lecture videos," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, p. 240, 2015.

[34] T.-J. K. P. Monserrat, S. Zhao, K. McGee, and A. V. Pandey, "Note-video: facilitating navigation of blackboard-style lecture videos," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 1139–1148.

[35] M. Kam, J. Wang, A. Iles, E. Tse, J. Chiu, D. Glaser, O. Tarshish, and J. Canny, "Livenotes: a system for cooperative and augmented note-taking in lectures," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2005, pp. 531–540.

[36] S. E. Garber, L. Milekic, A. Prakash, N. Moran, A. Di Lillo, and J. A. Storer, "Visual lecture summary using intensity correlation coefficient," in *Irish Machine Vision and Image Processing Conference Proceedings 2017*, 2017.

[37] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transactions on Graphics (TOG)*, vol. 2, no. 4, pp. 217–236, 1983.