J Sci Educ Technol
DOI 10.1007/s10956-016-9653-1

Indexed Captioned Searchable Videos: A Learning Companion for STEM Coursework

Tayfun Tuna[1] · **Jaspal Subhlok**[1] · **Lecia Barker**[2] · **Shishir Shah**[1] · **Olin Johnson**[1] · **Christopher Hovey**[2,3]

Abstract Videos of classroom lectures have proven to be a popular and versatile learning resource. A key shortcoming of the lecture video format is accessing the content of interest hidden in a video. This work meets this challenge with an advanced video framework featuring topical indexing, search, and captioning (ICS videos). Standard optical character recognition (OCR) technology was enhanced with image transformations for extraction of text from video frames to support indexing and search. The images and text on video frames is analyzed to divide lecture videos into topical segments. The ICS video player integrates indexing, search, and captioning in video playback providing instant access to the content of interest. This video framework has been used by more than 70 courses in a variety of STEM disciplines and assessed by more than 4000 students. Results presented from the surveys demonstrate the value of the videos as a learning resource and the role played by videos in a students learning process. Survey results also establish the value of indexing and search features in a video platform for education. This paper reports on the development and evaluation of ICS videos framework and over 5 years of usage experience in several STEM courses.

Keywords Assessment · Learning technologies · Lecture videos · OCR · Text segmentation · Video indexing · Video search · Video segmentation

✉ Tayfun Tuna
ttuna@uh.edu

Jaspal Subhlok
jaspal@uh.edu

Lecia Barker
lecia.barker@colorado.edu

Shishir Shah
sshah@central.uh.edu

Olin Johnson
johnson@cs.uh.edu

Christopher Hovey
hoveyc@colorado.edu

1 Department of Computer Science, University of Houston, Houston, TX 77204, USA

2 Department of Information Science, University of Colorado, Boulder, CO 80309, USA

3 Department of Sociology and Anthropology, Northeastern University, Boston, MA 02115, USA

Introduction

Video of classroom lectures is a versatile learning resource. It is often made available as additional material for a conventional course, as the core of distance learning coursework, or posted publicly for community learning or as reference material. Evidence of the popularity of lecture videos includes thousands of complete courses posted on portals such as MIT OpenCourseware and Apple's iTunes University. At the University of Houston, video lectures have been widely used for over a decade to enhance STEM coursework. Typically, Tablet PCs, which allow free mixing of prepared (PowerPoint) viewgraphs with hand annotations and illustrations, are employed for teaching and simultaneous recording of lectures. Advantages include excellent resolution, as the video consists of PC screen shots, and low video production cost, as no camera or operator is needed. The videos typically include whatever the professor is projecting on the screen (e.g., PowerPoint slides, animations, annotations, formulas, algorithms, or drawings) and the instructor's voice. Prior research has established that the recorded lectures are a powerful learning resource (Abowd 2000; Brotherton and Abowd 2004; Chandra 2007; Johnston et al. 2013; Nashash and Gunn 2003; Odhabi and Nicks-McCaleb 2011).

A major weakness of the video format is the inability to quickly access the content of interest in a video lecture. This is intuitively clear and also established by our surveys and student interviews (Barker et al. 2014). This paper reports on ICS videos: videos enhanced with indexing, captioning, and search capability that are designed for quick access to video content. Indexing adds logical index points, each in the form of a snapshot representing a video segment that can be accessed directly; captioning adds the transcript of the video lecture in a separate panel; and search enables identification of video segments that match a keyword provided by the user.

Significant research and integration challenges were addressed in the course of developing the ICS video player. Indexing and search features are based on the text displayed in the video. Enhancements to off-the-shelf OCR techniques were developed for efficient and accurate identification of text on video frames. New algorithms based on text analysis were developed to divide a video into segments that cover similar material. Finally, a user interface that integrates video playback with indexing, search, and captions was developed and refined with feedback from students using the system.

The ICS video framework has been used by dozens of STEM courses and 1000s of students at the University of Houston, a large public university system. Results are reported from extensive assessment of ICS video usage carried out with student surveys and instructor interviews. The research presented clearly demonstrates that (i) PC-based video lectures are a very valuable student resource, (ii) the framework developed to enhance videos with indexing and search features is efficient, effective, and a significant improvement over the state-of-the-art, and (iii) indexing and search capability significantly enhance the value of lecture videos. The video framework developed is freely available to academic institutions.

This paper is organized as follows. Section 3 presents the design and features of the ICS videos system, which is the centerpiece of this project. Section 4 describes how keywords are identified in video lectures. Section 5 describes the indexing process and the methodology developed for dividing a video lecture into segments representing different topics. Section 6 reports the execution performance of the proposed system. Finally, Sect. 7 describes the real-world usage of the ICS videos system and presents survey-based results on the overall value of ICS videos as well as the value of indexing and search features.

Related Work

We discuss the research related to the two major aspects of the work presented in this paper separately.

Usage and Value of Videos

Automatic capturing of video of class lectures and conference presentations, followed by their online presentation and distribution, has been deployed for many years (Abowd 2000; Ahanger and Little 1996; Ma et al. 1998; Tobagi 1995). These videos are mostly recorded by cameras installed in the lecture/presentation rooms (Bianchi 2004).

Previous research on student and faculty use and perceptions has found that lecture videos can be very advantageous to students. Students generally report using videos when they are available and find them beneficial to their learning (Abowd 2000; Brotherton and Abowd 2004; Chandra 2007; Johnston et al. 2013; Nashash and Gunn 2003; Odhabi and Nicks-McCaleb 2011). Video usage can improve student performance, participation, and course satisfaction (Brandsteidl 2012; Defranceschi and Ronchetti 2011; Lancaster et al. 2011; Dickson et al. 2012; Bannon et al. 2011; Toppin 2011; Traphagan et al. 2010). While some studies show that attendance rates can drop with video usage (Johnston et al. 2013; Traphagan et al. 2010; Bell et al. 2001), most studies demonstrate that attendance remains the same, or in some cases, even improves, with video usage (Brotherton and Abowd 2004; Brandsteidl 2012; Defranceschi and Ronchetti 2011; Toppin 2011; Briggs 2007; Chandra 2011; Copley 2004; Hew 2009; Nast et al. 2009). In a proper implementation in classroom, the faculty member draws attention to videos as a learning resource, gives instruction on how to access them, uses them throughout a semester, and uses material from the videos in assignments and examinations. These are important for adoption and value of videos, and variation in implementation may explain the difference in results in the studies.

A key to the success of lecture videos is that this format provides flexibility for students to review material multiple times and/or at their own pace until they have a firm grasp on the material (Johnston et al. 2013; Nashash and Gunn 2003; Dickson et al. 2012; Toppin 2011; Traphagan et al. 2010). While other reviewing platforms such as professor's notes or PowerPoint slides offer similar flexibility to positive effect (Babb and Ross 2009; Grabe 2005; Grabe and Christopherson 2005), the combination of verbal explanation, class discussion and follow-up questions, and visual information, help students retain more information than text-only resources (Nashash and Gunn 2003; Traphagan et al. 2010).

Video Indexing and Search

As the number of videos increase, attempts have been made to index the videos, create digital libraries, and overall improve the recording and access to videos. The approaches to indexing the content of a video lecture are in the form of manual indexing (Joukov and Chiueh 2003; Young 2012), semiautomatic indexing (Ma et al. 1998; Chaisorn et al. 2009) and automatic indexing (Bianchi 2004; Arman et al. 1993; Nagasaka and Tanaka 1992; Otsuji and Tonomura 1993).

Project Lectern II (Joukov and Chiueh 2003) employs touch-sensitive screen technology to build a digital desk to transparently capture classroom lecturing activity, manually edit lectures, and automatically upload to a Web server for viewing. Coursera (Young 2012), an interactive learning system, employs segmented videos which are manually edited. Hypervideo (Ma et al. 1998) allows a user to navigate between video chunks using manually generated annotations and hyperlinks.

Automatic indexing involves the detection of key frames or labels that indicate change of content in a video. A multitude of methods have been developed that use low-level image properties such as color and texture, to group contiguous video segments automatically (Hampapur et al. 1995; Hanger 1995; Mo et al. 2005), while lacking the ability to provide semantic indexing. We employ similar techniques as a preprocessing step for detecting the slide changes in lecture videos, before proceeding to topic-based indexing.

Various methods have been developed that use screen text or speech text for content-based video retrieval, semantic multimedia retrieval, and meta-data generation (Lin et al. 2004; Nandzik et al. 2013; Yang et al. 2013; Biswas et al. 2015; Monserrat et al. 2013; Tuna et al. 2015). Talkminer (Adcock et al. 2010) provides a keyword search inside the lecture video which is similar to our approach. To support indexing as well as keyword search inside a video, text extraction from videos based on OCR tools has been employed (Lienhart and Effelsberg 1998; Merler and Kender 2009). The approach taken in this paper employs a suite of image transformations to improve the quality of OCR results, which is inspired by (Merler and Kender 2009). Extraction and ranking of keywords from both OCR and automatic speech recognition (ASR) methods is discussed in (Yang et al. 2013; Biswas et al. 2015). Comparing the speech text segments for similarity to determine the topic boundaries is studied in (Lin et al. 2004) employing a semiautomatic dictionary-based approach to identify important features for comparison. The indexing algorithms employed in this paper are based on cosine text similarity. An important contribution of this paper is integration of automatic indexing into a custom video presentation system with an emphasis on usability and ease of access.

In summary, the key features of the work presented in this paper that differentiate it from most previous work are (i) automation of topic-based segmentation and keyword search of videos employing image analysis, OCR enhanced with image transformations, and text analysis, (ii) focus on recordings from tablet PC screens, and (iii) large-scale evaluation with student surveys.

ICS Video Portal

We describe the objectives, design, and implementation of the ICS video portal that includes the ICS video player as well as the user interface to upload, play, and search videos.

Objectives

The reason for developing a custom ICS video player was that the commercial state-of-the-art video players were not up to the task for ICS video project goals. Following were the key design objectives:

Video Streaming and Browsing Interface: Lecturers should be able to upload videos and share the links with the viewers so that they can browse and play. Categorization of lectures can be done by the subject, course, department, semester, or lecturer.

Automatic Indexing: The videos should be automatically divided into logical segments. These video segments should be represented in an intuitive way so that they are easy to access and navigate for the users.

Video Search: Users should be able to search for the occurrences of a keyword inside a video as well as across a library of videos. The search results should be presented such that the existence and frequency of keyword matches is intuitive and clear.

Caption Support: The player should be able to display captions and users should be able to browse the transcript. The captions may be provided by ASR (automatic speech recognition) tools or developed manually.

Ubiquitous Software: The player should work across common devices (pc, Mac, mobile phones), operating systems, and Internet browsers.

Features

We discuss the main features of the ICS video player and the user interface to explain how the design objectives were achieved.

ICS Video Player

Video playback panel, keyword search box, index panel, and caption panel are the main visible components of the ICS video player. A snapshot of the player highlighting the key features is shown in Fig. 1.

Video Playback Panel: The playback window shows the video being played. Captions are displayed when available. The bottom part of the player has a panel with the timeline indicating the progress of the video playback. Users can skip to any part of the video by selecting the desired position in the timeline.

Index Display: An index panel is situated at the bottom of the player, as shown in Fig. 1. Index points are listed horizontally in the index panel and with a screenshot of the video at that point of time, along with the corresponding location on the timeline indicated in *mm:ss* format. Users can select any index point from the list, and the video starts playing from the selected index point. When the mouse cursor is placed over an index point image, the image is magnified to provide a better visual display of the segment represented by the index point.

Keyword Search: A search box is located above the index points panel. In order to execute a search, a user types the keywords in the search box. All index segments that contain the search keyword are displayed as results, while the index segments not containing the keyword are grayed out. The matching keyword, along with the number of matches in the segment, is also displayed below the corresponding index point image. The search feature is also illustrated in Fig. 1 where 5 video segments represented by index points are presented as matches for the search term *insp*. Clicking on the search images displayed will start the playback from that point in the video.

Captions: Current caption is displayed as an overlay at the bottom of the main video player. The transcript of the video lecture divided by timestamps is displayed in a separate caption window on the right side of the player. The transcripts scrolls automatically and the current caption is highlighted. Users can also scroll back and forth to see the part of the transcript earlier or later than the current location of the playback. Clicking on a sentence in the transcript restarts the video at that point in the timeline.

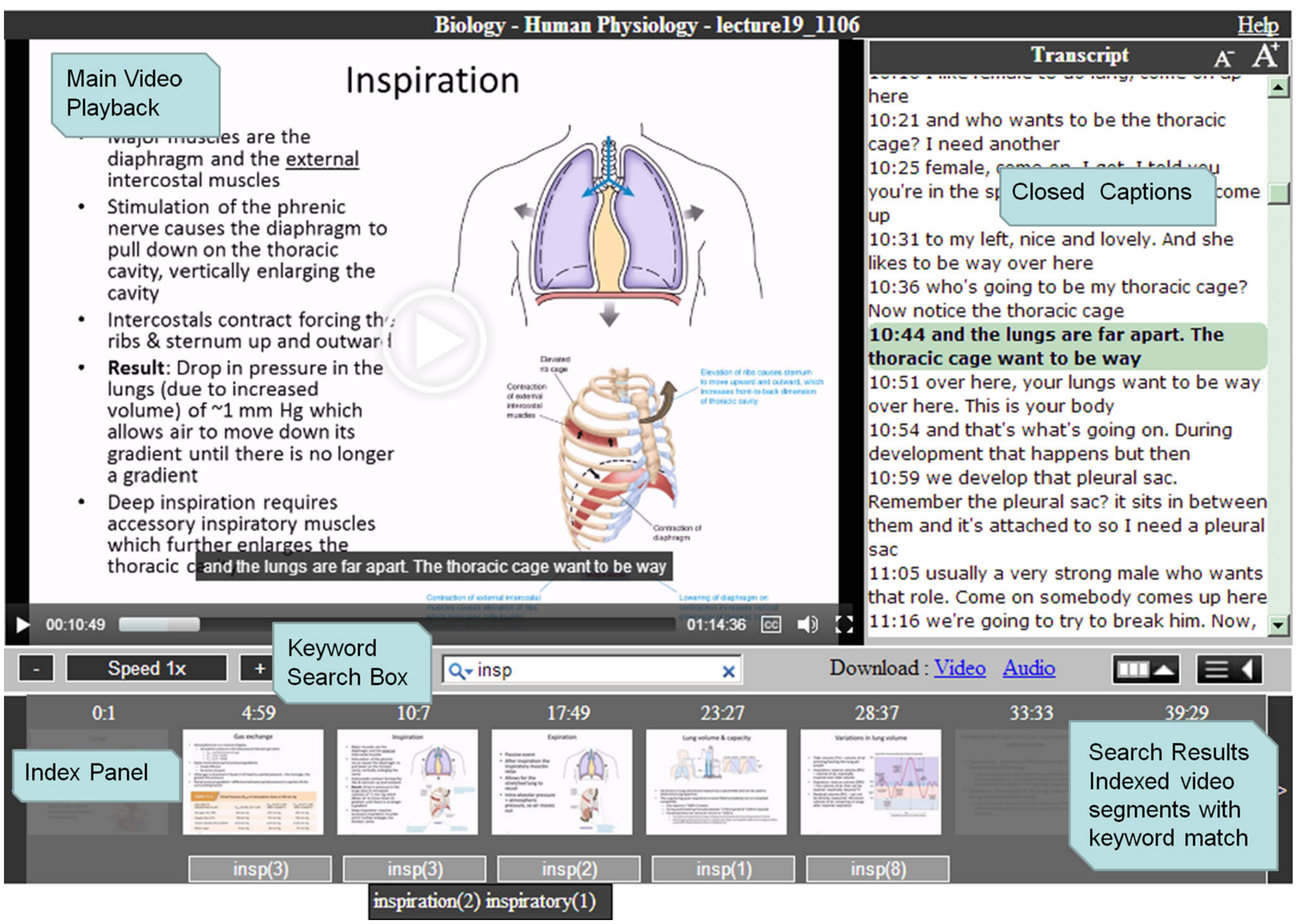


Fig. 1 A snapshot of the ICS video player

Convenience Features: Speed buttons can be used to decrease/increase the speed of the video playback. We expect that slowing down the video will help students catch difficult phrases. Speeding up the video can help a student proceed quickly to the point of interest. There is an option to download the video or just the audio. Size of the text in the transcript panel can be increased or decreased by the buttons A− and A+. Shortcut keys are also available to pause/play or to move to a different time in the video playback. The index panel, captions on the main video player, and the caption window can be turned on or off by clicking the appropriate buttons on the player.

Video Library Interface

The main functions of the ICS videos library interface are achieved through the video upload page, browse videos page, and search videos page illustrated in Figs. 2 and 3.

Video Upload: An instructor uploads videos from a video upload page as illustrated in Fig. 2a. Videos are categorized under department, semester, and course names. A video file is uploaded along with a caption file (optional).

Browse Videos: Videos for a particular offering of a course in a particular semester can be browsed. Figure 2b shows the 10 videos available for the *Digital Image Processing* course.

Video Library Search: Keywords can be searched across all videos available in the videos database. The results can be restricted to a department, a course, a semester, or a combination of these, as illustrated in Fig. 3a. In this example, the keyword *cell* is being searched in all departments and all courses for the Fall 2012 semester. The results show that *cell* exists in six lectures from Biology and Computer Science departments, in Human Physiology and Digital Image Processing courses. It shows the partial matches and the frequencies of the keyword matches. When a user clicks on a video link under *Lecture Description* column, it will direct the user to the search results in the corresponding video. This is illustrated in Fig. 3.

Implementation

The key technical challenges in implementing the ICS video player are automatic identification of search terms, automatic indexing, and generation of captions. Search keywords are determined by adopting and enhancing OCR technologies with a suite of image transformations to identify text in video frames. The framework for identification of text in video frames is described in Sect. 4. Index points are determined by an analysis of text patterns in video frames. The indexing algorithms employed are described in Sect. 5. The ICS video framework currently assumes that a caption file is provided by the user. Combining state-of-the-art speech recognition technologies with a crowdsourced caption editor has been addressed in related work (Deshpande et al. 2014).

Workflow for the ICS video framework is depicted in Fig. 4. Once a video is uploaded, the images on the video are extracted selectively. Text in the images is identified with OCR and image enhancements. The text in the images is then analyzed by the indexing module. The keywords

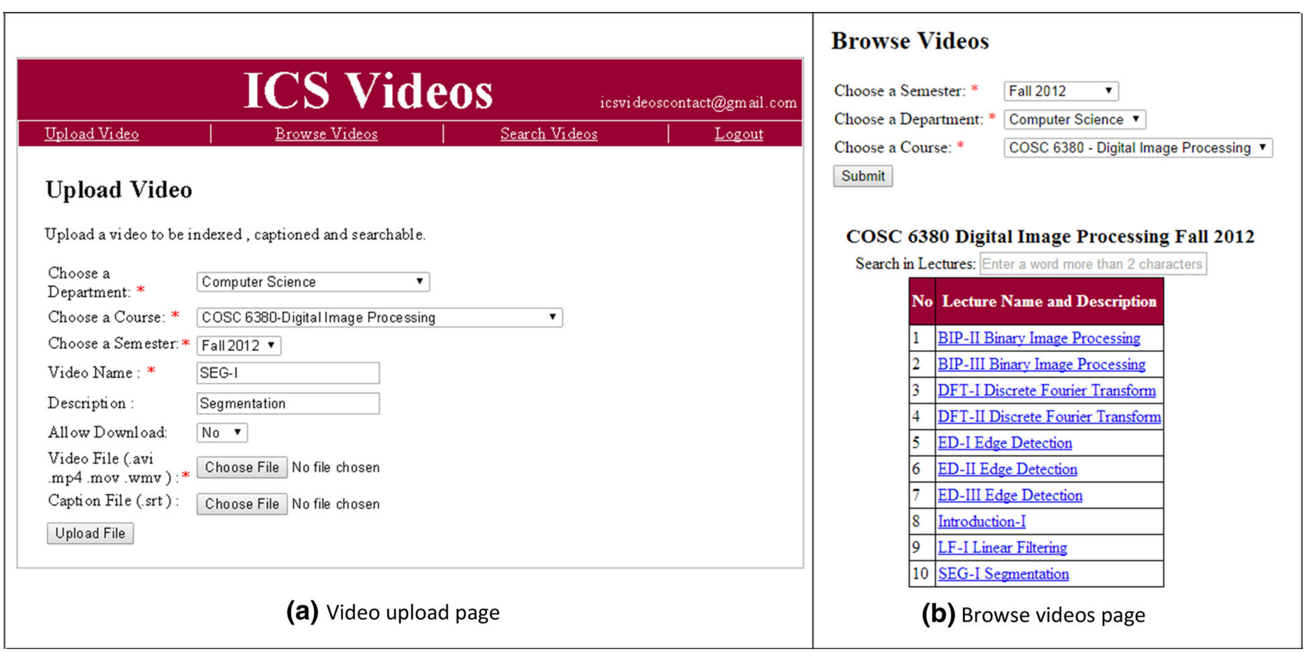
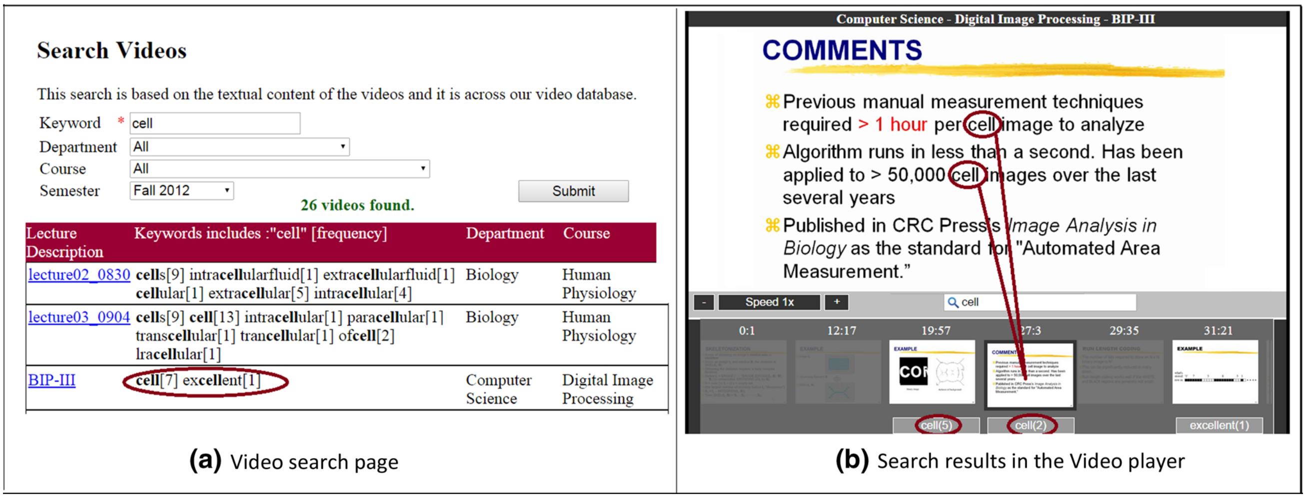


Fig. 2 ICS videos interfaces: upload and browse videos page



(a) Video search page **(b)** Search results in the Video player

Fig. 3 ICS videos interfaces: search videos page and video playback

and their location, along with the index points, are stored in the video player database. Content and time location of captions are also included in the database. The video player accesses the database on demand to support indexing, search, and captions for users.

ICS video streaming interface and player are entirely built on top of open-source technology; the software modules and versions currently employed are listed here. The Apache 2.4.9 Web Server is used for video streaming. PHP version 5.2.4 is used for interfaces. ICS player is an HTML5 video player built by the MediaElement.js. It is capable of playing standard H.264 encoded streaming media on HTML 5 compatible devices including PC, Mac, tablet, iPad, and most smartphones. The player uses two different video formats, *MP4* and *WEBM* to ensure compatibility with different browsers: Chrome, Firefox, Internet Explorer, Opera and Safari. MySQL version 5.5 is used

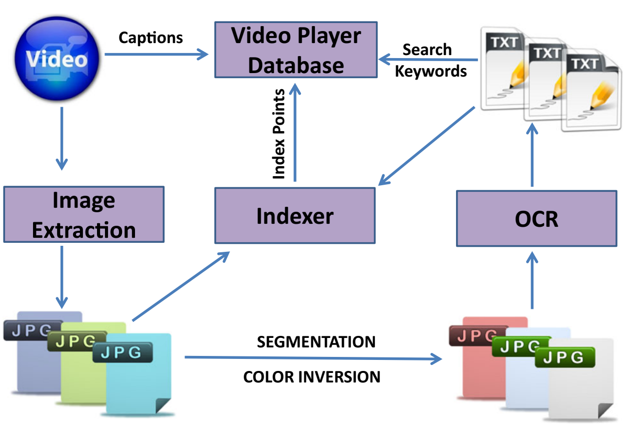


Fig. 4 ICS videos framework: First images corresponding to unique frames in the video are extracted. Images are enhanced for OCR text extraction which provides the search keywords. Video indexing uses the extracted images and text to create index points. The results are stored in a database for the ICS video player

as database and database updates are done in the background using AJAX(asynchronous JavaScript and XML). JQuery (version: 1.4.1), a JavaScript library, is used to traverse the HTML document and to make AJAX requests.

Keyword Search

Keyword search and video indexing require that the text contained in the video frames be identified. In this section, the methodology for recognizing text in a video frame is presented. Clearly, it is not necessary to process every frame in a video as consecutive video frames typically have identical text. Selection of frames for text recognition is part of our methodology for identification of index points discussed in Sect. 5.

Recognition of text in a video frame can be accomplished by the use of optical character recognition (OCR) tools, an approach investigated in (Merler and Kender 2009). We analyzed a suite of OCR tools for their effectiveness in recognizing text in video frames. The following tools were selected for a comprehensive evaluation: *GOCR*, an open-source program available under the GNU Public License, *Tesseract* developed at Hewlett Packard Labs and now managed and improved by Google, and MODI (Microsoft Office Document Imaging) toolset. We discovered that OCR tools generally have limited effectiveness at recognizing text in the presence of (1) certain combinations of text and background colors and shades, (2) text mingled with colorful shapes, and (3) small and exotic fonts.

Image Enhancements

To increase the detection efficiency of text in video frames, we investigated the use of several simple image processing techniques for image enhancement (IE) prior to the application of OCR tools. IE operations that were effective in enhancing text recognition included *segmentation* of text followed by enlargement with interpolation and *color inversion*. Additional details on application of IE techniques to improve OCR performance are available in (Tuna et al. 2011).

Segmentation

Segmentation of text involves a sequence of steps to define and extract the text regions in an image for improved OCR, as outlined in Fig. 5. Following is the sequence of steps in text segmentation.

1. *Binarization* converts the color image to a binary black and white image by using simple image statistics-based thresholding. Threshold is calculated as the sum of weighted pixel values divided by the sum of weights. An example binarized image is shown in Fig. 5b.
2. *Dilation* allows for expansion of separate objects that can result in merging of objects in close proximity, by eliminating small *holes* between them. We employ it for grouping the text characters and for identifying text regions in the image.
   The image after the dilation steps is shown in Fig. 5c.
3. *Edge Detection* attempts to connect incomplete borders of text regions after dilation. We employ an edge detection algorithm to complete the construction of text segments as illustrated in Fig. 5d. Several edge detection algorithms are available, and we chose the Sobel operator, which is one of the most commonly used edge detectors in image processing (Shrivakshan and Chandrasekar 2012).
4. *Blob Extraction* can extract standalone objects in an image. Blob extraction is used to detect the location of text segments in the dilated image as shown in Fig. 5d. The connected components labeling algorithm was employed for blob extraction (Fisher et al. 2003). In an extracted blob, one would expect more blobs; however, they were filtered using the following two criteria. If a blob contains other blobs, or if the $blobwidth/blobheight < 1.5$, we do not extract it. The text we want to detect is at least two characters long; hence, the width is always expected to be more than the height. In Fig. 5e, the man's body is not extracted because of the threshold on the height-to-width ratio. In addition, very small size blobs are not included in the extracted regions.
5. *Resizing* involved enlargement with interpolation that is implemented for the segmented blocks. By this operation, small size text is enlarged to become visible to OCR engines. Resizing is illustrated in Fig. 5f.

Color Inversion

Color inversion is achieved by altering the RGB (red, green, and blue) values of images, aimed at increasing the contrast between the text and background. In image file formats such as BMP, JPEG, TGA, and TIFF that commonly use 24-bit RGB representations, color value for each pixel is encoded using 24 bits per pixel. Three 8-bit unsigned integers (0 through 255) represent the intensities of red, green, and blue, respectively. Inverting colors is basically altering the RGB values. When we invert an image in a classical way, we take the inverse RGB values. For example, the inverse of the color (1,0,100) is (255-1,255-0,255-100) = (254,255,155). In our approach, we expand this technique from 1 to 7 inversions shown in Fig. 6, where **R'** is referring to the 255-R value. OCR engines often give different results for the original and

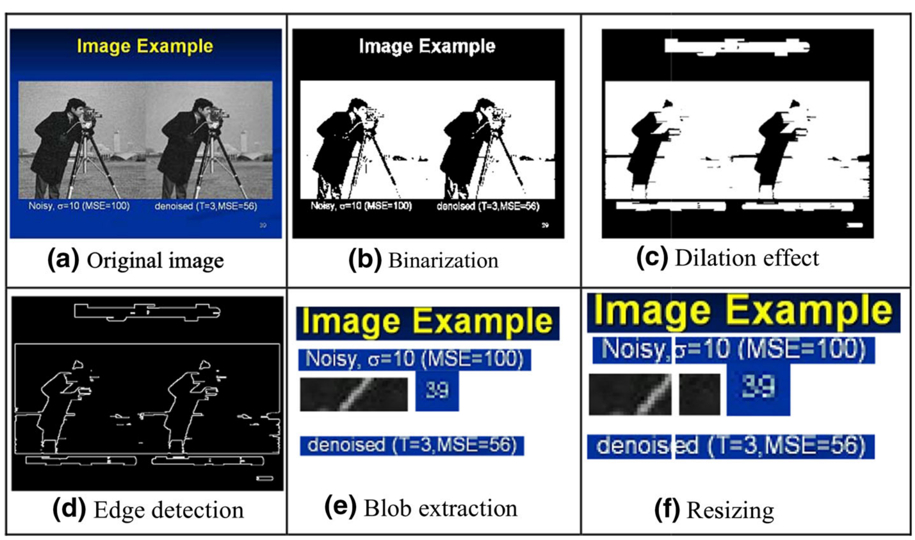


Fig. 5 Segmentation and enlargement of text

inverted images. Image enhancement procedures often lead to new text being recognized, but can also prevent the recognition of other text. Hence, OCR engines are applied to the original images as well as the inverted images and the union of the results is taken. An example of color inverted images is shown in Fig. 6.

Evaluation

To test the OCR tools and the impact of the image enhancement procedures, we evaluated 1387 different images that were selected by the indexer from 20 diverse videos. Images in these videos contain 20,007 unique words, 27,201 total words (of more than 1 character length) for a total of 144,613 characters. *Search accuracy* is defined as the number of detected unique words divided by the total number of unique words, whereas *false-positive ratio* is defined as number of falsely detected unique words divided by total number of unique words.

$$\text{Search Accuracy} = \frac{\text{Number of Correctly Detected Unique Words}}{\text{Total Number of Unique Words}} \quad (1)$$

$$\text{False Positive Ratio} = \frac{\text{Number of Falsely Detected Unique Words}}{\text{Total Number of Unique Words}} \quad (2)$$

Experimental results, presented in Fig. 7, show that the search accuracy of three distinct OCR engines, Tesseract, GOCR, and MODI, improved 8.8 % on average, with image enhancements. The maximum accuracy obtained by applying all OCR engines with image enhancements was 97.1 %. Alternately stated, the miss rate was 8.9 % for the best single OCR engine, 5.2 % for all OCR tools combined, and 2.9 % for all OCR engines combined with image enhancement.

Image enhancement provided this accuracy improvement, but increased the processing time significantly, partly because the OCR engines have to be applied on the original and the enhanced images. Nonetheless, the overall processing time remains modest for a typical video. On average, it is in the range of 2–3 min for an hour long video as detailed in Sect. 6. Image enhancement also

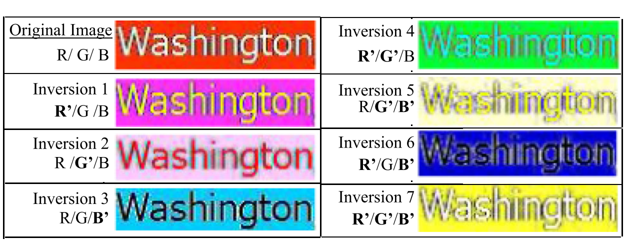


Fig. 6 Inversion example: original image and color inverted images

significantly increases the false positives detected by OCR engines, i.e., more words were detected that were not actually present in the video as shown in Fig. 7. This often happens when an OCR engine misses a character in a word, leading to false identification of a different word. Since the main aim of the text recognition is to let the user find words of interest, the extra words resulting from false positives are unlikely to diminish the search functionality in a meaningful way.

Indexing

Textbooks are organized by chapters and sections based on topics and subtopics. A reader typically does not read a textbook from the beginning to the end, but uses it to read about certain topics or review some concepts. A reader can immediately find a chapter in the book from the table of contents or find locations where a topic is discussed based on keywords from the index at the end of a textbook. In contrast, accessing the content of interest in a lecture video is not easy because there are no table of contents or index sections. Often the only way to find a topic of interest in a video is by scrolling the video from the beginning, which can be time-consuming and frustrating, especially for long videos. Video indexing aims to overcome this challenge by dividing a lecture video into segments that contain different topics or subtopics. The beginning of each segment is called an *index point* that is visually represented by the image at the beginning of that time segment. A user can visually see the location of various topics and subtopics with these index point images and can easily access the content of interest or switch between topics by clicking on the index point images.

This section presents the process of automatically dividing the video into segments by discovering where the topics or subtopics have changed. In the first step, *transition points*, i.e., places where the scene in the video changes, are identified. Transition points typically relate to viewgraph changes in a lecture. Next, the text on the transition points is gathered with methods discussed in Sect. 4. In the final step, the text on the transition points is analyzed to identify the *index points* that relate to topic changes.

Identification of Transition Points

Detection of transition points in a video is based primarily based on identifying scene changes, and secondarily on text changes, on the video frames. Corresponding pixels in successive frames are considered to be different if they differ by a minimum RGB threshold when the RGB values of the pixels are compared. Successive frames constitute a

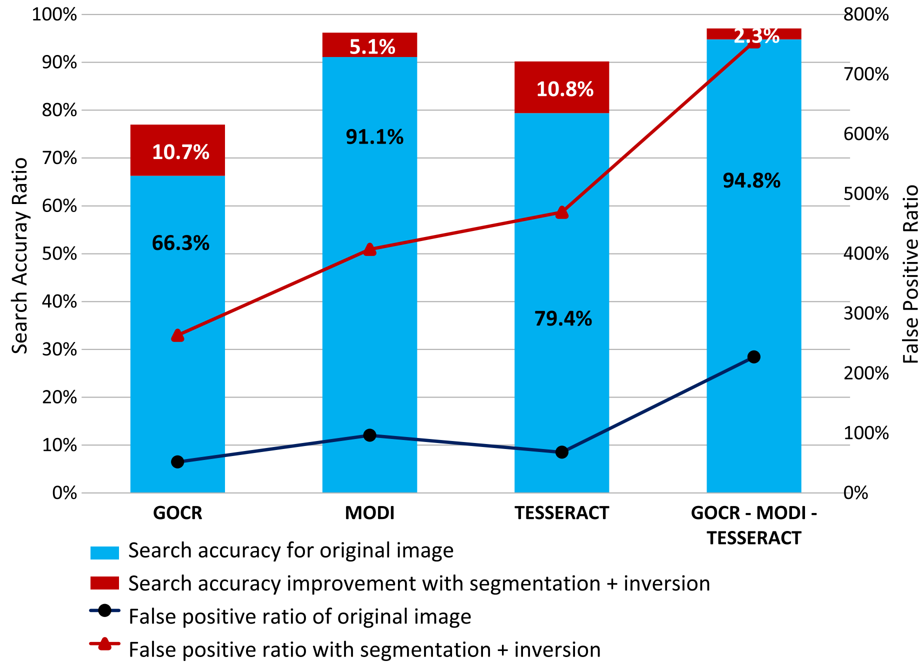


Fig. 7 Search accuracy and false-positive ratios for OCR engines with original image and enhanced image: Image enhancements increase the search accuracy, but it significantly increases the false positives

transition point if the fraction of pixels that are different based on the RGB criteria exceeds a minimum threshold, which we refer to as the transition point threshold. The reason for using thresholds to identify transition points is that frames corresponding to the same scene in practice (e.g., exactly the same viewgraph) also have minor differences in the RGB spectrum that must be ignored to avoid false transition points. The threshold values are chosen empirically. A value of 10 % was selected for both RGB threshold and transition point threshold for the system deployed in this paper. Figure 8 illustrates a transition point in a sequence of video frames.

ICS lecture videos framework is primarily designed for screencasts with video image changing only periodically when the viewgraph(typically power point slide) changes. For such videos, identifying scene changes based on RGB pixel difference is sufficient to identify the transition points. But some videos also have webcam recordings, video clips inside the video, animations, or other dynamic content which led to a large number of false transition points with the RGB pixel difference approach as screen image changes continuously. To handle such scenarios, the text content on the video frame is also examined; if two frames have identical text on them, they are merged into a single transition segment even if the difference in the images exceeds the RGB pixel thresholds.

Comparing pairs of all successive frames in a video is rather inefficient and time-consuming. We apply two methods to speed up this process: Sampling and Binary Search.

Sampling: In a video lecture, scene transitions are relatively infrequent. With this optimization, instead of comparing successive frames, the current frame is compared to a frame that is one jumping interval ahead. If the two frames are identical, then in all likelihood, the frames between them are also identical. Hence, the search marker is moved up by a jumping interval and the frames between them are not processed. If the current frame and the frame one jumping interval ahead are different, then the frames between them are compared sequentially starting from the current frame to identify all transition points. Essentially, in a long sequence of identical frames, most comparisons between successive frames are skipped.

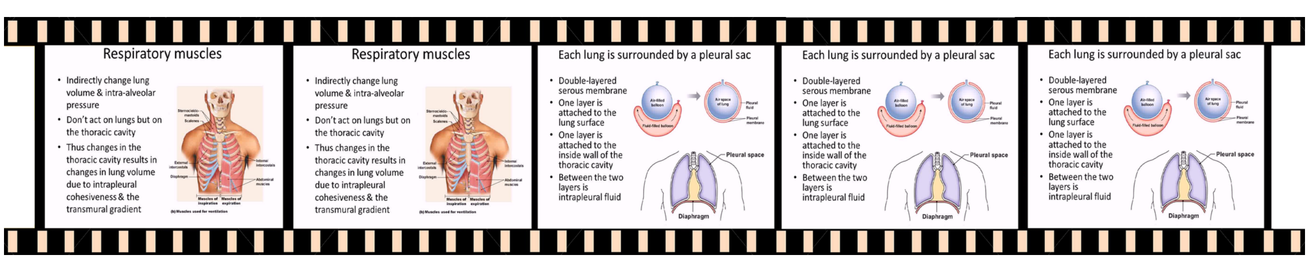


Fig. 8 Transition point in a video: Third frame represents a transition point

Binary Search: This procedure starts by splitting the video into two segments—from the first frame to the middle frame and from the middle frame to the last frame. For each of these segments, the first and the last frame are compared. If they are identical, then it is assumed that the segment contains no transition points. Otherwise, the segment is again subdivided into two segments and the procedure is applied recursively. When a segment size of one is reached, a transition point is identified. We note that some transition points can be missed in some pathological cases due to these optimizations. However, the cost of losing a few transition points is outweighed by the significant benefit of improved execution time as detailed in (Tuna et al. 2012, 2011).

Identification of Index Points

A lecture video often has a large number of transition points; over 100 transition points is not unusual. The ultimate goal is to identify a smaller number of index points that are related to topic changes. The text on transition frames is identified with OCR tools and image transformations as discussed in Sect. 4. We now present the text analysis techniques employed to discover topic changes in a video lecture.

Data: List of transition segments represented by a vector of words:
T= $t_1, t_2, t_3..t_n$
Required number of index points: k
Result: k index points that are a subset of T
repeat
  t_s=transition segment with smallest duration;
  if $cossim(t_s, t_{s+1}) > cossim(t_s, t_{s-1})$ **then**
    merge(t_s, t_{s+1});
  else
    merge(t_{s-1}, t_s);
  end
until *number of segments in* $T == k$;

Algorithm 1: Outline of text-based indexing algorithm

The core idea of text-based indexing is that different topics are represented by different groups of words. Comparing the frequencies of different words in blocks of text establishes how similar they are in content and topic. Intuitively, a video is split into different topical segments at the point where the mix of words being used in video frames changes significantly. But how do we compare the similarity of two blocks of text? While many different text similarity metrics have been discussed in the literature, we used *cosine similarity*, a well-known and proven metric used in information retrieval and text mining (Huang 2008; Tata and Patel 2007; Ye 2011). It is a measure of similarity between two vectors, calculated by the dot product of the vectors divided by the product of their norms as shown by the formula below.

$$cos(\theta) = \frac{A.B}{\|A\|.\|B\|} = \frac{\sum_{i=1}^{n} A_i.B_i}{\sqrt[2]{\sum_{i=1}^{n} A_i^2}.\sqrt[2]{\sum_{i=1}^{n} B_i^2}} \quad (3)$$

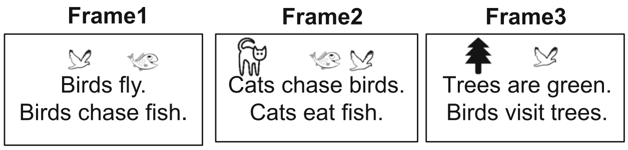
An example of text similarity calculation is depicted in Fig. 9. Video frames are shown in Fig. 9a, and the word frequency vector for each of these frames is represented in Fig. 9b. The cosine similarity between the vectors representing adjacent frames is computed as follows: CosineSimilarity(frame 1, frame 2) = 0.57 and CosineSimilarity(frame 2, frame 3) = 0.14. This matches the intuitive judgement that frame 1 and frame 2 are more similar to each other than frame 2 and frame 3. The implication is that any topic change inside this sequence should start with frame 3.

We sketch the overall text-based indexing procedure as Algorithm 1. The inputs are transition segments and the desired number of index points, and the output is a list of index points. The algorithm repeatedly merges the smallest transition segment in the video to the previous or the next segment based on text similarity until the number of segments equals the desired number of index points. The desired number of index points is given as a parameter.

Evaluation

Text-based indexing algorithm identifies the topical index points among the numerous transition points. The accuracy of the text-based indexing algorithm was evaluated with 25 lecture videos, primarily from Computer Science, Biology and Geology departments.[1]

[1] This set of videos is not the same as the videos used for evaluating keyword search in Sect. 4, but there is a large overlap. The reason for using slightly different sets of videos was the lack of availability of the instructors who provided ground truth for some of the videos at the time the evaluation was conducted.



(a) Example sequence of text frames

Word/Frame	Frame 1	Frame 2	Frame 3
birds	2	1	1
cats	0	2	0
chase	1	1	0
fish	1	1	0
fly	1	0	0
green	0	0	1
trees	0	0	2
visit	0	0	1

(b) Word frequency vectors for each frame

Fig. 9 A sequence for frames and their frequency vectors computed to determine similarity between the text on the frames

The common approach to evaluate index point selection in such a scenario would be a confusion matrix where the list of index points in the ground truth is compared against the index points reported by the automatic indexing algorithm. A major difficulty in evaluating the algorithm is that the ground truth, i.e., the true set of index points, is often not uniquely defined. It is very challenging even for an expert to decide whether a transition point is the start of a subtopic or not. Ground truth is also subjective; we have verified that different subject experts come up with significantly different set of index points for a lecture. Our experience led us to an unconventional approach to evaluation, where the creator of each lecture video (normally the instructor teaching the course) was asked to rate every transition point on its appropriateness to be an index point, instead of providing a list of index points. Based on the extent to which a transition point represented a change in the topic, they were marked on the following scale: Definitely Index Point (+2), Probably Index Point (+1), Probably Not Index Point (−1), and Definitely Not Index Point (−2). For the set of videos used for evaluation, out of 1628 transition points, the instructors assigned 31 as "Definitely Index Point," 49 "Probably Index Point," 69 "Probably Not Index Point," and 1379 "Definitely Not Index Point."

The output of the indexing algorithms is binary, i.e., each transition point is determined to be an index point (1) or not an index point (−1). The quality of the set of index points identified by an automatic indexing algorithm is determined as follows. Suppose the ground truth for a transition point is "Definitely Index Point." Then if the algorithm correctly identifies it as an index point, +2 is scored, while if it is incorrectly identified as not an index

		Ground Truth			
Algorithm Output		Definitely Not IP -2	Probably Not IP -1	Probably IP +1	Definitely IP +2
	Not IP -1	+2	+1	-1	-2
	IP +1	-2	-1	+1	+2

Fig. 10 Video indexing scoring for different ground-truth values and algorithm results

point, then −2 is scored. Now suppose the ground truth for a transition point is "Probably Index Point." Then if the algorithm correctly identifies it as an index point, +1 is scored, while if it is incorrectly identified as not an index point, then −1 is scored. Scoring is similar for the segments "Definitely Not Index Point," and "Probably Not Index Point." The scoring mechanism is illustrated in Fig. 10.

The sum of all individual scores is added to determine the raw indexing score for a video that we label as the video indexing score(VIS).

Suppose the video lecture contains n transition points. Then if G_i and A_i are the ground-truth score and algorithm score of transition point i, then the overall video indexing score is represented as:

$$\mathrm{VIS} = \sum_{i=1}^{n} G_i * A_i \tag{4}$$

Finally, the (relative) accuracy score of an algorithm for a video is the video indexing score (VIS) computed for the video as a percentage of the theoretical maximum VIS score for the video corresponding to theoretically optimal indexing.

Figure 11 shows the mean, median, and quartiles for different indexing methods. The accuracy scores represent the average over 25 videos in our test set. The text-based indexing algorithm is compared to uniform indexing and random indexing. In random indexing, index points are randomly selected. Uniform indexing aims to divide a

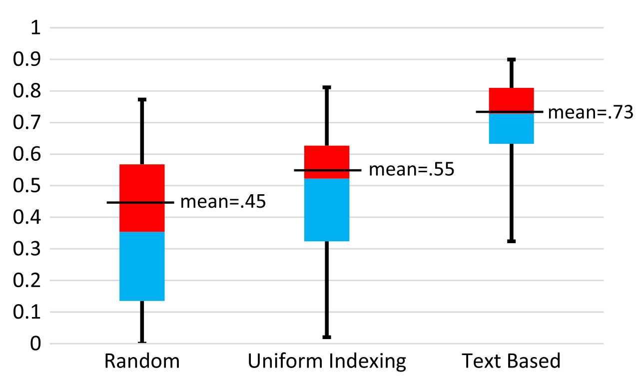


Fig. 11 Accuracy of different video indexing methods

video lecture into approximate uniformly distributed segments. It follows the steps in Algorithm 1 except that it does not use text information to compare frames but simply merges a segment with the smaller segment on the left or right. Figure 11 shows that the accuracy of the text-based indexing algorithm(73 %) is significantly higher than the accuracy with uniform indexing (55 %) and random indexing (45 %). In other words, the text-based algorithm was more accurate than the random indexing ($t = 9.064, df = 24,\ P\ < .001$), having an average accuracy rate that was 35.7 % higher. Similarly, the average accuracy rate of the text-based algorithm was 22.4 % higher than that of the uniform indexing algorithm ($t = 7.522, df = 24, P < .001$).

Execution Performance

We summarize the processing time and workload for analysis of videos to support the ICS features. The major steps in the overall process of indexing a video are (i) the extraction of video frames for analysis, (ii) pixelwise comparison of video frames for identifying transition points, (iii) extraction of text from video frames, and (iv) running of the text-based indexing procedures. The actual time to process a video is naturally dependent on the machine (or server) running the ICS framework, the resolution and format of the video, and to some extent on the content of the video. Our objective here is to get a broad idea of the execution characteristics of this framework.

Figure 12 presents the execution time data as the average over 25 videos on a typical server machine: Intel Core 2 Quad 2.4 GHz CPU, 8 GB Ram, 64-Bit Windows 7 Operating System. The overall processing of videos took just over 3 min (192 s) for an hour of video. Bulk of the time goes for frame extraction (38 %) and frame comparison for identifying transition points (55 %). Text extraction with OCR takes around 7 % of time, while the running time of the indexing algorithm is minuscule (0.52 %).

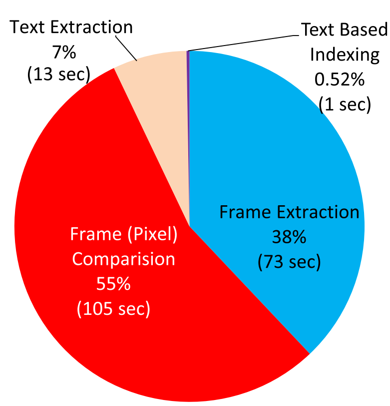


Fig. 12 Average time (seconds) for processing an hour of video

The main significance of these results is that the execution capability needed to support ICS videos is low enough that the hardware cost for building a video infrastructure is modest; a typical server class machine can process hundreds of videos a day. Another implication is that since text extraction and indexing take only a small fraction of the overall time, more computationally expensive approaches may be feasible. In ongoing work, we are exploring the use of machine learning for video segmentation.

Assessment

The ICS video system has been used widely at University of Houston in science, technology, engineering, and mathematics (STEM) coursework for several years. Students are surveyed each semester about the lecture videos and specifically about the indexing, captioning, and search features. The surveys include questions regarding video access including the nature and frequency of use, need, value, and class preparation; education experience including expected grade, credit hours taken during the semester, hours spent studying per week, academic year; and extracurricular experience including hours worked to earn income, commute time, marital status, number of dependents, and demographic information. In this section, we present results that focus on the students' perceived value of videos and the features presented in this paper.

Table 1 captures the usage of videos over the period reported in this paper. A total of 1602 videos were uploaded in this period. We note that a lecture video is accessed on average around 42 times. The table shows that indexing, search, and speed up/down features all found significant usage. There are large differences in the usage of these features across lectures, and a detailed discussion of the usage analytics is beyond the scope of this paper. However, the data clearly show large usage of videos and the features developed in this work.

Table 1 ICS videos usage statistics for total 1602 videos uploaded between Fall 2012 and Spring 2016

	Total #of activity	Average #of activity per video
Video view	67566	42.2
Search in a video	21057	13.1
Search inside a class	3776	2.4
Search across video library	1417	0.9
Index click	58836	36.7
Speed up	27659	17.3
Speed down	18286	11.4

Surveys were conducted across several courses over a span of 5 years. There are 73 courses offerings in computer science, biology, geology, chemistry, mathematics, physics, and psychology. Surveys were administered online during the last 2 weeks of the semester. Invites were sent to students by their instructor via e-mail and contained a hyperlink to the online survey. In this section, we include result from 2010, 2011, and 2013. Assessment was not conducted in 2012. It should also be noted that the survey instrument was revised between each semester, so some results reported in this section do not include data from every year.

Value of Lecture Videos

This section presents results compiled from a total of 1167 students between 2010 ($n = 612$ students) and 2011 ($n = 555$ students). Table 2 shows results of the several survey questions regarding students' perceptions of how valuable they found the lecture videos. Items were scaled to reflect increasing ratings: The first four items were rated on a scale of 1 (*disagree strongly*) to 6 (*agree strongly*), and the final item was rated on a scale of 1 (*not at all important*) to 4 (*very important*). As the table shows, students overwhelmingly reported that the lecture videos were important and valuable for reviewing, studying for quizzes and examinations, and clarifying material that was not clear in the class. Students also overwhelmingly reported that videos were important in getting the grade they hoped to get in the class. Similar sentiments were expressed in written comments; one student wrote, "*Using the videos really helps get the grade I wanted. All biology department professors should do this.*"

To develop a deeper understanding of student perceptions, students were asked to report on how they used lecture videos. Figure 13 shows the results of a survey with multiple selection options. On average, students selected around 4 options (mean = 3.94, $S = 2.15$), though there was relatively high variability in how many responses students

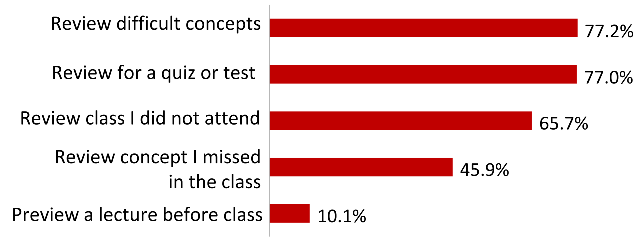


Fig. 13 Student-selected purpose of use ($N = 444$)

chose: Some students selected all options and others only selected one response. Of the 444 students who answered this battery, the most commonly reported use of lecture videos was to review difficult concepts, or concepts they did not understand during class (77.2 %). Slightly under two-thirds of the respondents (65.7 %) reported using recorded lecture videos to make up for a class they missed. Another 45.9 % used videos to review material they missed in a class they had attended, be it due to not being able to hear the lecturer, a brief distraction in class, or because their attention was limited by taking notes. For example, one student noted *"I would view the lecture once, but pause it and replay it constantly, to write down extra notes that I might have missed during the first viewing. This was extremely helpful to be able to do this"* Clearly, students made good use of the videos.

Similar to other studies of recorded lecture videos (Brotherton and Abowd 2004; Brandsteidl 2012; Defranceschi and Ronchetti 2011; Harley et al. 2003; Lancaster et al. 2011), Fig. 13 shows that student usage of the videos typically spiked before examinations: More than three-fourths of students (77.0 %) reported using lectures before a quiz or test. Around 10 % of students reported using the videos to preview a lecture before going to class; however, the significance of this number is not clear as videos were not always available before a class. Students were asked to rate the importance of lecture videos in comparison with other resources made available by faculty, including professors' lecture notes, students' own notes, and the

Table 2 Lecture video value ratings

Survey item	N	Mode	Mean	Std. Dev.
Lecture videos are useful for reviewing	905	6	5.63	.691
Lecture videos help me to clarify material that was not clear in class	887	6	5.45	.821
The lecture videos helped me to study for quizzes or tests	891	6	5.51	.830
Having access to lecture videos for this class is important to me	899	6	5.62	.767
How important was use of the lecture videos for this class getting the grade you hoped for?	885	4	3.51	.751

Mode 6 levels are *disagree strongly(1)*, *disagree(2)*, *disagree slightly(3)*, *agree slightly(4)*, *agree(5)*, *and agree strongly(6)*

Mode 4 levels are *not at all important(1)*, *slightly important(2)*, *somewhat important(3)*, and *very important(4)*

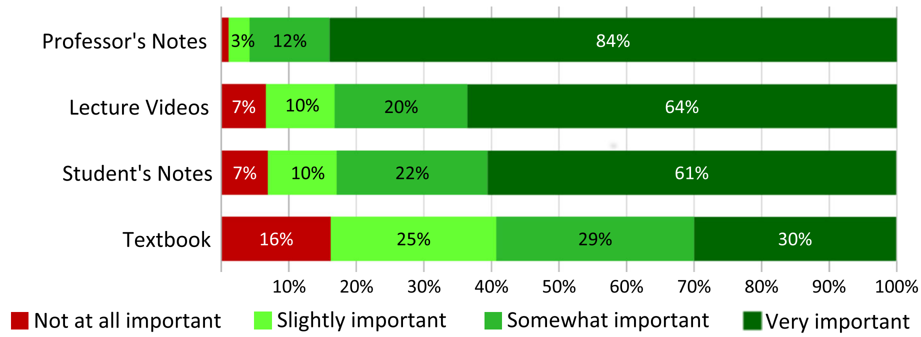


Fig. 14 Student ratings of studying resources

textbook assigned for each class. The results are shown in Fig. 14. In relation to getting the grade they wanted for the class, students gave the highest ratings for professors' lecture notes (84.0 % considered them to be *very important*), which adds support to earlier studies that have shown providing full or *complete* lecture notes can be beneficial to students (Babb and Ross 2009; Grabe 2005; Grabe and Christopherson 2005). Lecture videos received the second highest rating, with 63.6 % of students reporting that this resource was *very important* in getting the grade they wanted. Students own notes were slightly less valued, with 60.5 % of students giving a rating of *very important*. A low percentage of students, only 30 %, felt that textbooks were *very important* to their performance in the course, a finding echoed in a similar study by Evans (Evans 2008). Many students used a variety of sources to maximize learning; for example, one student reported, *"I actually really like that this class has lecture videos available. It's really helpful for me to go to class, then read the book about the lecture, and then watch the video lecture again when studying a week or so before the test. This method has been very successful for me and I've made A's on all the exams in this semester."*

Indexing

Figures 15 and 16 show the response of approximately 120 students from Spring 2013 and Fall 2013 semester to a forced-answer question about the usefulness and value of the indexing tool. In line with previous studies (Brandsteidl 2012; Dickson et al. 2012), our results confirmed that being able to quickly locate pertinent segments through search and index features can enhance the value and use of lecture videos.

Figure 15 shows that between 93 and 96 percent of respondents agreed that the index was helpful, the placement of index points in the video timeline was appropriate for the lectures, the layouts of the index images made the index feature easy to use, and the index points separated a lecture into logical segments. In this figure, *Disagree*

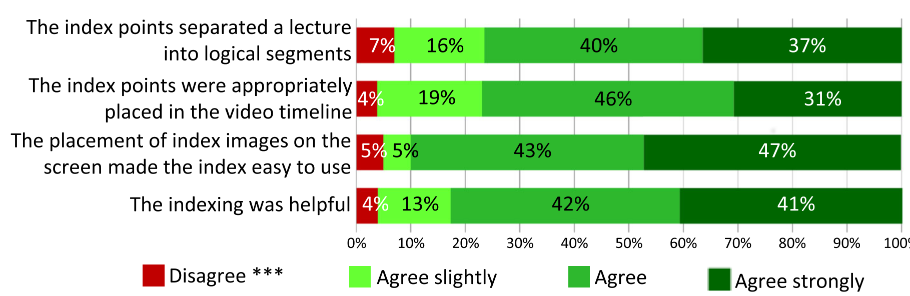


Fig. 15 Value of indexing

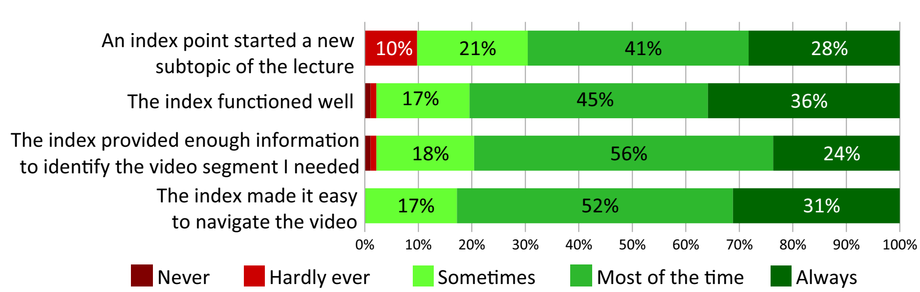


Fig. 16 Usability of indexing

Fig. 17 Value of search

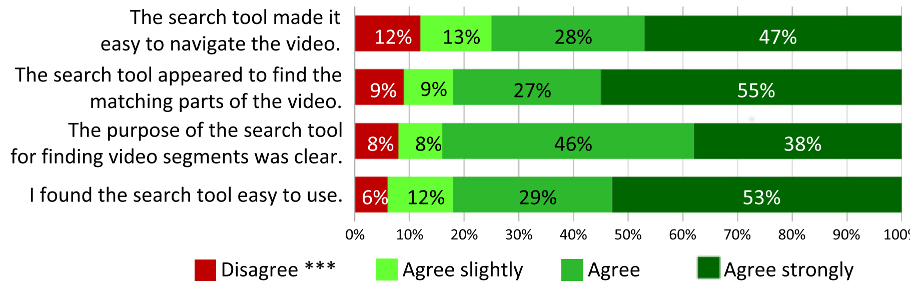
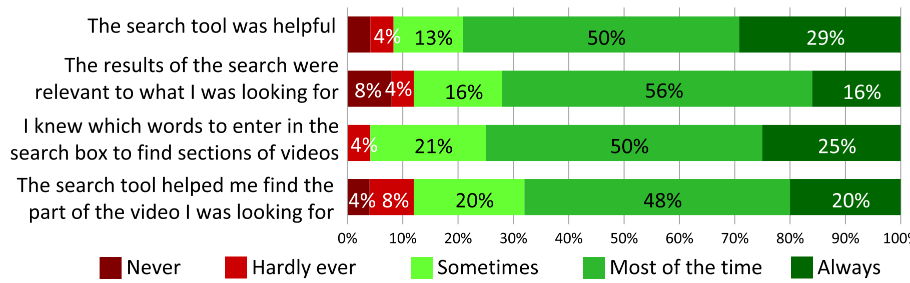


Fig. 18 Usability of search



strongly, Disagree, and *Disagree slightly* are merged to *Disagree**** due to the low number of responses.

Responses to additional questions on the value of indexing are presented in Fig. 16. Students are strongly supportive of the statements that the index feature functioned well, that the index points provided enough information to identify video segments of interest, and that the index made it easy to navigate the video. The statement that index points represented the start of a new subtopic had somewhat weaker support than the other assertions (10 % of students said *hardly ever*, 21 % said *sometimes,* 41 % said *most of the time*, and 28 % said always). This was not surprising given the difficulty of automating an accurate identification of new topics in a video previously discussed in Sect. 4. However, it is important to note that even imperfect indexing is perceived as very valuable by the students.

In open-ended comments, students reported several benefits from using the index including (a) saving time—for example one student wrote, "*I didn't have to wade through the rest of the lecture just to answer one question*"; (b) skipping through material the student was familiar with to get to the challenging sections; and (c) returning to a section of the lecture if an interruption occurred. For example, one student wrote, "*Sometimes I would have to pause the lecture to take care of other responsibilities that I had to attend to, and when I was ready to come back to the lecture I'd pick up exactly where I was at. It was great*!" Another student said "*The indexing feature, in my opinion, is one of the best parts regarding this video player. It separated the lecture into reasonably sized sections and made it easy to know where to pick a lecture back up if I had to stop watching for a while.*"

Keyword Search

Figures 17 and 18 show the responses to the questions on the keyword search. The response rate was low for this set of questions; only 39 students responded. We believe there are several reasons for this. Many students may not see a need for using the keyword search feature as indexing allows navigation of topics inside a video. Index points are clearly visible when the ICS video player is active and navigation only requires clicking on the index snapshots. In order to utilize the search feature, the user needs to identify the search box and identify and type the search keywords. Also, the exact functionality of the search box may not be obvious to some students and earlier versions of the player had the search box located in a corner that was not conspicuous.

Nonetheless, of the 39 students who used it, 94 % of respondents reported that the search feature was easy to use, and 81 % thought the results were appear to be true. While 70 % felt that feature helped them find the part of the video they intended to find most of the time, only 75 % reported that they usually knew which words to enter into the search box to find the segment of video they wanted. We speculate that if instructors can increase their students familiarity with the content and proper keywords, the usage of search to find the intended video clips should increase.

Additional results in Fig. 18 show that, at least some of the times, 88–96 % of students (depending on the item) found the search tool helpful, found that the results of the search feature were relevant to what they were looking, knew which words to enter in the search box to find sections, and thought the search tool helped them to find the part of the video they were looking for.

In summary, the results show that keyword search was found to be very valuable by the students who used it. However, underutilization remains a problem because the feature may not be needed by all students. Additional training and familiarity with the search feature is needed for wider adoption.

Faculty Perspectives

Input from instructors was sought frequently through annual meetings, interviews, and electronic communication. Instructor perception and experience was a key driver in enhancing the ICS videos framework.

Overall, the instructors reported a very positive experience. They reported that the framework was easy to work with and offered significant benefit to the students. We also received strongly positive assessment of ICS feature's including indexing, search, and speed control.

The complaints mainly consisted of hardware and compatibility issues, such as noise, low microphone level, and incompatibility with some mobile devices and browsers. Additional suggestions included navigation through blocks of index points, better location of controls such as the search bar, improving the visual appeal of the user interface, integration with audio, and adding more details on index points. The player is being continually enhanced, and many of these suggestions/complaints have already been addressed.

In 2015, two faculty members were formally interviewed about their experience with ICS videos: one who taught a traditional lecture format, and another a flipped class. Sample quotes, edited for clarity, are included below.[2]

On the novel features of the player

"Search feature is phenomenal. I think the two features that are absolutely great are the ability to speed up the video and the keyword search. When the students are first introduced to it, they find it absolutely amazing. And what I find is that most students initially watch the whole lecture all the way through which is really kind of a time waster. When you're going to the videos to look up something, you really want to just get to your information and the search feature directs you to those keywords."

On the positive and negative aspects of the videos

"Most positive effect of having that recorded lecture available to the student is that it allows them to go back after the lecture is done to compare their notes versus what they've actually heard in class. And what's important about that is that it gives the student a little bit of freedom to actually engage in the classroom discussion as opposed to furiously writing down notes and serving as a stenographer for my lecture. So as a tool to help prepare for the material it is a highly effective supplement to the lecture. The downside is that some students see the lecture videos as a substitute for the lectures. So some students, generally not the top students but those who find themselves on the bottom levels of the class, tend to stop coming to class. My own anecdotal evidence is that the students find that, when they start skipping class and just start relying on the videos, grades go down. So they generally start to come back to class."

On the overall value of the toolset

"It is a wonderful product. I mean faculty that don't use this are doing it to the student's detriment. It is the same thing as saying that I refuse to use blackboard or refuse to use power point in the classroom. I mean any tool that you can use that can improve a student's ability to learn and to retain information should be something that you incorporated in the classroom."

Discussion and Future Work

The work presented in this paper demonstrates with large-scale experimentation that students find substantial value in recorded lecture videos as a companion to traditional classroom lectures. Students primarily used videos for the purpose of review and clarification of difficult concepts and not as a substitute for face-to-face contact. Lecture videos are a dynamic and versatile resource that can promote learning and, in combination with other pedagogical innovations like active learning, allow faculty to invert their classrooms and try new methods of fulfilling their roles as educators.

The paper also demonstrates that indexing and search are critical features that enhance the value of videos. Automatic methods for enabling those, while far from perfect, do provide a practical and valuable addition to video presentation.

It is our hope that the educational vendors and developers who provide the infrastructure to make lecture videos possible pursue the following directions:

- Enable instructors to customize videos with easy interfaces that allow indexing and other basic editing.
- Simplify the recording and production process, from the start of a recording to the end when students start viewing.
- Enhance and automate captioning to make video material more useful to deaf and non-native language speaking students.
- Make infrastructure for ICS or similar video technology widely available to STEM instructors.

[2] Full interview videos are available in http://icsvideos.uh.edu/interviews2015/

– Integrate video infrastructure with common course management platforms like BlackBoard and Moodle.

We also expect that future work from this group and others will provide clear results on the impact of videos on learning under various models of education.

Conclusion

This paper reports on technologies developed for indexed, captioned, and searchable videos and their usage for STEM coursework. We demonstrated that automated indexing and search frameworks are effective and efficient. The ICS videos framework underwent large-scale deployment and assessment and was a success in all respects. The videos were judged by students to be very valuable and employed for diverse purposes such as clarifying and reviewing material and preparing for examinations. Indexing and search features were considered very helpful and easy to use. This work points to a new and innovative direction for effective use of videos in STEM coursework. The ICS videos framework is currently being used by dozens of courses, primarily at University of Houston. It is freely available to educational institutions.[3]

References

Abowd GD (2000) Classroom 2000: an experiment with the instrumentation of a living educational environment. IBM Syst J 38:508–530

Adcock J, Cooper M, Denoue L, Pirsiavash H, Rowe LA (2010) Talkminer: a lecture webcast search engine. In: Proceedings of the international conference on Multimedia, ACM, New York, NY, USA, MM ’10, pp 241–250

Ahanger G, Little TD (1996) A survey of technologies for parsing and indexing digital video. J Vis Commun Image Represent 7:28–43

Arman F, Hsu A, Chiu MY (1993) Image processing on compressed data for large video databases. In: MULTIMEDIA ’93: Proceedings of the first ACM international conference on Multimedia, ACM Press, New York, NY, USA, pp 267–272. http://doi.acm.org/10.1145/166266.166297

Babb KA, Ross C (2009) The timing of online lecture slide availability and its effect on attendance, participation, and exam performance. Comput Educ 52(4):868–881

Bannon BWO, Lubke JK, Beard JL, Britt VG (2011) Using podcasts to replace lecture: effects on student achievement. Comput Educ 57(3):1885–1892

Barker L, Hovey CL, Subhlok J, Tuna T (2014) Student perceptions of indexed, searchable videos of faculty lectures. In: Proceedings of the 44th annual frontiers in education conference (FIE), Madrid, Spain

Bell T, Cockburn A, McKenzie B, Vargo J (2001) Flexible delivery damaging to learning? Lessons learned from the canterbury digital lectures project. In: Proceedings of world conference on educational multimedia, hypermedia and telecommunications, pp 117–122

Bianchi M (2004) Automatic video production of lectures using an intelligent and aware environment. In: Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia, ACM, New York, NY, USA, MUM ’04, pp 117–123

Biswas A, Gandhi A, Deshmukh O (2015) Mmtoc: A multimodal method for table of content creation in educational videos. In: Proceedings of the 23rd ACM international conference on multimedia, ACM, New York, NY, USA, MM ’15, pp 621–630. doi:10.1145/2733373.2806253

Brandsteidl M, Mayerhofer T, Seidl M, Huemer C (2012) Replacing traditional classroom lectures with lecture videos: an experience report. In: Proceedings of the 8th edition of the educators symposium, New York, NY, USA pp 21–27

Briggs LL (2007) Classroom capture: lecture recording system draws devotees at temple. Campus Technology Digital Magazine

Brotherton JA, Abowd GD (2004) Lessons learned from eclass: assessing automated capture and access in the classroom. ACM Trans Comput-Hum Interact 11(2):121–155

Chaisorn L, Manders C, Rahardja S (2009) Video retrieval evolution of video segmentation, indexing and search. Computer Science and Information Technology pp 16–20

Chandra S (2007) Lecture video capture for the masses. In: Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education, ACM, New York, NY, USA, ITiCSE ’07, pp 276–280. doi:10.1145/1268784.1268864

Chandra S (2011) Experiences in personal lecture video capture. IEEE Trans Learn Technol 4(3):261–274

Copley J (2004) Audio and video podcasts of lectures for campus-based students: production and evaluation of student use. Innov Educ Teach Int 44(4):387–399

Defranceschi A, Ronchetti M (2011) Video-lectures in a traditional mathematics course on itunes u: Students feedback. 2011 14th international conference on interactive collaborative learning (ICL) pp 219–224

Deshpande R, Tuna T, Subhlok J, Barker L (2014) A crowdsourcing caption editor for educational videos. In: Proceedings of the 44th annual frontiers in education conference (FIE), Madrid, Spain

Dickson PE, Warshow DI, Goebel AC, Roache CC, Adrion WR (2012) Student reactions to classroom lecture capture. In: Proceedings of the 17th ACM annual conference on innovation and technology in computer science education, New York, NY, USA pp 144–149

Evans C (2008) The effectiveness of m-learning in the form of podcast revision lectures in higher education. Comput Educ 50(2):491–498

Fisher R, Perkins S, Walker A, Wolfart E (2003) Connected component labeling

Grabe M (2005) Voluntary use of online lecture notes: correlates of note use and note use as an alternative to class attendance. ComputEduc 44(4):409–421

Grabe M, Christopherson K (2005) Evaluating the advantages and disadvantages of providing lecture notes: The role of internet technology as a delivery system and research tool. Internet High Educ 8(4):291298

Hampapur A, Jain R, Weymouth TE (1995) Production model based digital video segmentation. Multimedia Tools Appl 1(1):9–46. doi:10.1007/BF01261224

Hanger A (1995) A survey of technologies for parsing and indexing digital video. Technical Report TR-11-01-95, Boston University

Harley D, Henke J, Lawrence S, McMartin F, Maher M, Gawlik M, Muller P (2003) Costs, culture, and complexity: an analysis of technology enhancements in a large lecture course at UC berkeley. Tech. rep, UC Berkeley

[3] http://icsvideos.uh.edu

Hew KF (2009) Use of audio podcast in k-12 and higher education: a review of research topics and methodologies. Educ Technol Res Dev 57(3):121–155

Huang A (2008) Similarity Measures for Text Document Clustering. In: Holland J, Nicholas A, Brignoli D (eds) New Zealand Computer Science Research Student Conference, pp 49–56

Johnston ANB, Massa H, Burne THJ (2013) Digital lecture recording: a cautionary tale. Nurse Educ Pract 13(1):40–47

Joukov N, Chiueh T (2003) Lectern ii: a multimedia lecture capturing and editing system. Int Conf Multimed Expo 2:681–684

Lancaster JW, McQueeney ML, Amburgh JAV (2011) Online lecture delivery paired with in class problem-based learning Does it enhance student learning? Curr Pharm Teach Learn 3(1):23–29

Lienhart R, Effelsberg W (1998) automatic text segmentation and text recognition for video indexing. Multimed Syst 8:69–81

Lin M, Nunamaker J JF, Chau M, Chen H (2004) Segmentation of lecture videos based on text: a method combining multiple linguistic features. In: Proceedings of the 37th annual Hawaii international conference on system sciences, p 9

Ma WH, Lee YJ, Du DC, McCahill M (1998) Video-based hypermedia for education-on-demand. MultiMed IEEE 5(1):72–83. doi:10.1109/93.664744

Merler M, Kender JR (2009) Semantic keyword extraction via adaptive text binarization of unstructured unsourced video. In: Proceedings of the 16th IEEE international conference on Image processing, IEEE Press, Piscataway, NJ, USA, ICIP'09, pp 261–264

Mo H, Yamagishi F, Ide I, Katayama N, Satoh S, Sakauchi M (2005) Key image extraction from a news video archive for visualizing its semantic structure. Adv Multimed Info Process 3331:650–657

Monserrat TJKP, Zhao S, McGee K, Pandey AV (2013) Notevideo: Facilitating navigation of blackboard-style lecture videos. In: Proceedings of the SIGCHI conference on human factors in computing systems, ACM, New York, NY, USA, CHI '13, pp 1139–1148. doi:10.1145/2470654.2466147

Nagasaka A, Tanaka Y (1992) Automatic video indexing and full-video search for object appearances. In: Proceedings of the IFIP TC2/WG 2.6 second working conference on visual database systems II, North-Holland Publishing Co., pp 113–127

Nandzik J, Litz B, Flores-Herr N, Lhden A, Konya I, Baum D, Bergholz A, Schnfu D, Fey C, Osterhoff J (2013) CONTENTUS technologies for next generation multimedia libraries. Multimed Tools Appl 63(2):287–329

Nashash HA, Gunn C (2003) Lecture capture in engineering classes: bridging gaps and enhancing learning. Educ Technol Soc 16(1):69–78

Nast A, Schäfer-Hesterberg G, Zielke H, Sterry W, Rzany B (2009) Online lectures for students in dermatology: a replacement for traditional teaching or a valuable addition? J Eur Acad Dermatol Venereol 23(9):1039–1043

Odhabi H, Nicks-McCaleb L (2011) Video recording lectures: student and professor perspectives. Br J Educ 42(2):327–336

Otsuji K, Tonomura Y (1993) Projection detecting filter for video cut detection. In: MULTIMEDIA '93: Proceedings of the first ACM international conference on Multimedia, ACM Press, New York, NY, USA, pp 251–257. doi:10.1145/166266.166295

Shrivakshan G, Chandrasekar C (2012) A comparison of various edge detection techniques used in image processing. IJCSI Int J Comput Sci Issu 9(5):269–276

Tata S, Patel JM (2007) Estimating the selectivity of tf-idf based cosine similarity predicates. SIGMOD Rec 36(2):7–12

Tobagi F (1995) Distance learning with digital video. MultiMedia IEEE 2(1):90–93. doi:10.1109/93.368609

Toppin IN (2011) Video lecture capture (VLC) system: a comparison of student versus faculty perceptions. Educ Inf Technol 16(4):383–393

Traphagan T, Kucsera JV, Kishi K (2010) Impact of class lecture webcasting on attendance and learning. Educ Technol Res Dev 58(1):19–37

Tuna T, Subhlok J, Shah S (2011) Indexing and keyword search to ease navigation in lecture videos. In: Applied Imagery Pattern Recognition (AIPR), pp 1–8

Tuna T, Subhlok J, Barker L, Varghese V, Johnson O, Shah S (2012) Development and evaluation of indexed captioned searchable videos for stem coursework. In: Proceedings of the 43rd SIGCSE technical symposium on computer science education, ACM, pp 129–134

Tuna T, Joshi M, Varghese V, Deshpande R, Subhlok J, Verma R (2015) Topic based segmentation of classroom videos. In: Proceedings of the 45th annual frontiers in education conference (FIE), El Paso, Texas

Yang H, Grnewald F, Bauer M, Meinel C (2013) Lecture video browsing using multimodal information resources. In: Wang JF, Lau R (eds) Advances in web-based learning ICWL, vol 8167. Springer, Berlin, pp 204–213

Ye J (2011) Cosine similarity measures for intuitionistic fuzzy sets and their applications. Math Comput Modell 53(12):91–97

Young JR (2012) Inside the coursera contract: how an upstart company might profit from free courses. Chron High Educ 19(7):2012