Indexed Captioned Searchable Videos

Jaspal Subhlok

University of Houston
Department of Computer Science
Outline

- Technology and Learning (& MOOCs)
- Motivation for Employing & Enhancing videos
- How does it work: Demo
- Building It: The Technology
- Using It: Deployment & Lessons from Surveys
- Wrap Up

[The opinions in this talk are those of the speaker and do not represent the official view of Department of Computer Science or any funding agency]
Technologies Influence Education

1. (1480s) The Printing Press -> Textbook
   - + Fears of Professors were misplaced (settled in ~20 years)
   - + for education quality (not sure about memory capacity)

2. Computer, Viewgraphs/Powerpoint Lectures
   - + Professor & Student convenience
   - Influence on education less clear

3. Clickers, Video, Social Media, Online xxx, Tablets, Mobile Devices, Wikipedia,....
   - Not clear where this is heading
   - Model of lecturing+homeworks+exams intact so far!
What About MOOCs

Presented as an alternative model, but

- Motivated by cost reduction, but no convincing business model or integration with classrooms
- Evolution of distance learning and Evolution of textbooks
- Very valuable in many scenarios but overhyped
- Distraction from other technological developments with potential for impact
Drivers of Change?

- Learning literature says “students not learning much” (critical thinking, reasoning, writing)
- Financial squeeze as governments reduce/stop funding education. More like a business.

Education enterprise will change with maturing of new technologies and approaches (e.g. “Inverted/Flipped” classrooms)

- Events like shutting/scaling down of (smaller) universities more likely than ever!
- Combination of technology and external pressures
Research project to **enhance the value of video as a learning resource** employing automated (where possible) **Indexing, Search and Captioning.**
Motivation for ICS Videos Project

- University of Houston has been a leader in supporting video for coursework (NSM IT)
  - Tablet PCs to teach and record lectures and make them available online as study material

Surveys showed videos are a powerful, versatile learning resource
Survey Results - 1

- **N=2,349** taken from 43 sections in biology, computer science, geology, chemistry, and mathematics/physics between 2009 to 2011.
- Course sizes varied from 8 students to >300.
- Each professor posted ~25 videos per semester.

*These surveys are recent but the basic results have been similar even before this project started.*

*Survey instrument details are often skipped in this talk to save time – available in papers.*
Students Watch Videos

84% used lecture videos at least once

81% watched the entire video, not only a part they needed

47% watched a single video more than once
Reasons Students Used Videos

- Review for Test or Assignment: 77%
- Review Difficult Material: 76%
- Make Up for Absence: 70%
- Review Something I Didn't Hear in Class: 36%
- Review Concepts I Missed for Other Reasons: 20%
"Having access to lecture videos for this class is important to me."

"Lecture videos are useful for reviewing."

"The lecture videos helped me to study for quizzes or tests."

"Lecture videos help me to clarify material that was not clear in class."

<table>
<thead>
<tr>
<th></th>
<th>Agree strongly</th>
<th>Agree</th>
<th>Agree slightly</th>
<th>Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Having access...&quot;</td>
<td>72%</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Lecture videos...&quot;</td>
<td>72%</td>
<td>23%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;The lecture v...&quot;</td>
<td>68%</td>
<td>22%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Lecture videos...&quot;</td>
<td>59%</td>
<td>32%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Comparison with other Resources

- Professor's notes: 86%
- Lecture Videos: 66%
- Student's Notes: 60%
- Textbook*: 30%

- Very important
- Somewhat important
- Slightly important
- Not at all important

Department of Computer Science
Comparison with other Resources

- **Professor's notes**: 86%
- **Lecture Videos**: 66%
- **Student's Notes**: 60%
- **Textbook***: 29%

Legend:
- Very important
- Somewhat important
- Slightly important
- Not at all important

Department of Computer Science
Videos are valuable but:

- Key shortcoming of video format is the inability to quickly access content of interest
  - Loud and clear in surveys and interviews (not shown)
  - Students wants answers to questions for review, not watch an hour long video!

Goal: ICS video player with advanced Indexing, Captioning, and Search

A Project to bring clear and present benefits to students with cool computing research on the way!
Okay so I need 3 more volunteers for today to give your powerpoint.

presentation. Let’s see you are Shannon, are you Shannon?

Jasmine, okay so you’re in the third. Jasmine Scott.

What’s your name now? Your last name is?

okay I got 2, I need one more? We’ll go with these two okay. First one is going to be Jasmine Scott here.

So Jasminen is going to talk to us about TranSwitch: Engines for Global Connectivity. Hello everybody, I’m Jasmine. The company I chose was TranSwitch.

Basically its a company, they make integrated circuits and intellectual property solutions.
Building It – Challenges

**Search:** Keyword search Inside video
- OCR on video frames
- Semantic search

**Indexing:** Divide a lecture into topic segments
- Identifying topic changes
- Images, text, or audio?

**Captioning:**
- Speech recognition ineffective for classroom videos
Keyword Search

- Keyword Search requires text detection in video frames
- Can be accomplished by OCR tools

- Accuracy on lecture video images?
Images from Lecture Videos

- OCR: good for scanned images
- Lecture video images with colorful different layouts

---

**Link Layer: Introduction**

Some terminology:
- hosts and routers are **nodes**
- communication channels that connect adjacent nodes along communication path are **links**
  - wired links
  - wireless links
  - LANs
- layer-2 packet is a **frame**, encapsulates **datagram**

**data-link layer** has responsibility of transferring **datagram** from one node to adjacent node over a link

---

**Testcross**

- How can you determine the genotype of a dominant phenotype?:
  - Purple-flowered pea plant is either **PP** or **Pp**; cross with a homozygous recessive (**pp**) white-flowered plant:

  - If **PP**
    - **Sperm**: \( P \) \( p \)
    - **Eggs**: \( P \) \( p \)
    - **All purple**
    - or
  - If **Pp**
    - **Sperm**: \( P \) \( p \)
    - **Eggs**: \( Pp \) \( Pp \) \( pp \) \( pp \)
    - **1/4 purple; 3/4 white**

- By definition, the **testcross** is used to determine the genotype of an organism expressing a dominant phenotype by breeding with a recessive homozygote.
Question 3

Where did the story say that there was a statue raised in Mrs. Bethune’s honor?

- Washington, D.C.
- Miami, Florida
- Mayesville, South Carolina
Segmentation of Text Regions

a) Original image
b) Binarization
c) Dilation effect
d) Edge detection
e) Blob extraction
f) Enlargement

Color Inversion

Original Image  
R / G / B

Inversion 1  
255-R / G / B

Inversion 2  
R / 255-G / B

Inversion 3  
R / G / 255-B

Inversion 4  
255-R / 255-G / B

Inversion 5  
R / 255-G / 255-B

Inversion 6  
255-R / G / 255-B

Inversion 7  
255-R / 255-G / 255-B
Impact of Image Enhancement (IE)

Accuracy = %age of words detected correctly:

<table>
<thead>
<tr>
<th>Without IE</th>
<th>IE improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modi-Gocr-Tesseract</td>
<td>94.8%</td>
</tr>
<tr>
<td>Tesseract</td>
<td>79.4%</td>
</tr>
<tr>
<td>Goocr</td>
<td>66.3%</td>
</tr>
<tr>
<td>Modi</td>
<td>91.1%</td>
</tr>
</tbody>
</table>

Many false positives – *not a major issue for Search*
Value of Search

Results from Survey of Student Users
The survey results are as follows:

- **Agree strongly**: The search tool made it easy to navigate the video.
- **Agree**: The search tool appeared to find the matching parts of the video.
- **Agree slightly**: The purpose of the search tool for finding video segments was clear.
- **Disagree**: I found the search tool easy to use.
- **Disagree slightly**
- **Disagree strongly**

The percentage of students who agreed with each statement are shown in the diagram.
The search tool was helpful.
The results of the search were relevant to what I was looking for.
I knew which words to enter in the search box to find sections of videos.
The search tool helped me find the part of the video I was looking for.
Objective: Split video into meaningful, topical segments

1. Identify *Transition Points (TP)* where video scene changes (image-difference)

2. Identify subset of Transition Points that represent Topic changes!
Methods for Video Indexing

Split video in equal intervals of time
- Lazy reference method

Split video based on image difference
- Does not work well (J. Li MS thesis)

Split video based on text difference

Indexing by Machine Learning
Text Based Indexing Algorithm

Input:
- A list of transition points
- Required number of index points

Output:
- List of index points

Repeat:
1. Select transition point/segment with smallest duration
2. Merge it left or right neighbor based on text similarity

Until
Reach required number of index points
Uniform Indexing Algorithm

desired # index=5

TPs: 1 2 3 4 5 6 7 8

IPs: 1, 2, 4, 5, 7
**Text Similarity Metric: Cosine**

**Frame 1**
- Birds fly.
- Birds chase fish.

**Frame 2**
- Cats chase birds.
- Cats eat fish.

**Frame 3**
- Dogs swim.
- Dogs chase cats.

<table>
<thead>
<tr>
<th>Word/Frame</th>
<th>Frame1</th>
<th>Frame2</th>
<th>Frame3</th>
</tr>
</thead>
<tbody>
<tr>
<td>birds</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>cats</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>chase</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>dogs</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>fish</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>fly</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>swim</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\text{Cos(frame2,frame1)} = 0.80
\]

\[
\text{Cos(frame2,frame3)} = 0.57
\]
Video Indexing Experiment

- 25 different videos
  - Computer Science 19, Biology 3, Geology 3
  - 10 different course
  - Average of 75 minutes per video,
  - Total 30+ hours of video

- 3 Manual Ground Truths
  - 1700 transition points
Interface to Identify Ground Truth

3: Definitely Index Points,
2: Probably Index Point,
1: Probably Not an Index Point,
0: Definitely Not an Index Point

basin fill this segment we will examine the factors that come into play with respect to the architecture of basin fill these include the interactions between

basin fill
## Indexing Accuracy Metric

The table below shows the manual ground truths for different algorithm outputs and indexing accuracy scores:

<table>
<thead>
<tr>
<th>Algorithm Output</th>
<th>Definitely Not IP</th>
<th>Probably Not IP</th>
<th>Probably IP</th>
<th>Definitely IP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (Not IP)</td>
<td>(+2)</td>
<td>(+1)</td>
<td>(-1)</td>
<td>(-2)</td>
</tr>
<tr>
<td>1 (IP)</td>
<td>(-2)</td>
<td>(-1)</td>
<td>(+1)</td>
<td>(+2)</td>
</tr>
</tbody>
</table>

**TIS: Transition Indexing Score**

\[
\text{Video Indexing Score} = \sum_{i=1}^{n} (TIS_i)
\]

n: number of TP in a video
Indexing Algorithms Accuracy

Total Video Indexing Scores for 25 Videos

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>1300</td>
</tr>
<tr>
<td>Uniform</td>
<td>2100</td>
</tr>
<tr>
<td>Text Based</td>
<td>2300</td>
</tr>
<tr>
<td>Theoretical Max</td>
<td>3300</td>
</tr>
</tbody>
</table>
**Indexing Algorithms Accuracy**

**Indexing Accuracy for Total of 25 videos**

<table>
<thead>
<tr>
<th>Video Indexing Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>random</td>
<td>44%</td>
</tr>
<tr>
<td>uniform</td>
<td>67%</td>
</tr>
<tr>
<td>text based</td>
<td>73%</td>
</tr>
</tbody>
</table>

**Video Indexing Accuracy** = \[
\frac{\text{Video Indexing Score}}{\text{Theoretical Max Score}}
\]
Indexing by Machine Learning

- Promising but not fully validated! (Tuna’s Ph.D. thesis)
More on Indexing

- Employment of Speech-text with slide-text being studied (M. Joshi M.S. Thesis)
- Semi-automatic user driven indexing may be a useful option

Student Surveys of Value of Indexing
The index points separated a lecture into logical segments.
The index points were appropriately placed in the video timeline.
The placement of index images on the screen made the index easy to use.
The indexing was helpful.
An index point started a new subtopic of the lecture.
The index functioned well.
The index made it easy to navigate the video.
The index provided enough information to allow me to identify the video segment I needed.
ICS Videos: Captions

On screen Caption and Transcript panel
Captioning

- Captions (and transcript) are valuable for students

- Automatic speech recognition is inadequate to generate meaningful captions today

- Goal is to leverage ASR to generate captions efficiently
## Automatic Speech Recognition (ASR)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Lecture Speech Recognition</th>
<th>Dictation</th>
<th>Parroting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor A</td>
<td>71.40</td>
<td>89.93</td>
<td>94.76</td>
</tr>
<tr>
<td>Professor B</td>
<td>62.14</td>
<td>83.13</td>
<td>96.63</td>
</tr>
<tr>
<td>Professor C</td>
<td>70.80</td>
<td>83.03</td>
<td>96.1</td>
</tr>
<tr>
<td>Average</td>
<td>68.11</td>
<td>85.36</td>
<td>95.83</td>
</tr>
</tbody>
</table>

**Percentage Accuracy**

Results shown for the best of 3 ASR tools: YouTube, Windows, Dragon

**Dictation:** Participants read a prescribed text that is recorded

**Parroting:** A trained speaker repeats the text

Normal Lectures cannot be transcribed with ASR today
Motivation: A mechanism to provide captions for STEM classroom lectures.

Design Objectives/Challenges:

- **Teamwork:** Allow co-operative/crowdsourced captioning
- **Technical content + Poor audio quality:** Captions/transcript may not be obvious
- **Easy navigation:** for non专业人士.
## ICS Caption Editor

Welcome ruchal! Logout

### Legend
- Section currently being edited
- Caption is Complete
- Caption Needs a Review

### PlaySpeed
- Real

### 0 Completed!, 0 Requested for Review, 14 remaining

<table>
<thead>
<tr>
<th>Section No.</th>
<th>Start Time (mm:ss)</th>
<th>Caption Text</th>
<th>Save Changes</th>
<th>Status</th>
<th>Review Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edit Section 1</td>
<td>0:05</td>
<td>Let's talk about floating point today. We have seen the ways of representing numbers in binary so far</td>
<td></td>
<td>Needs a Review</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0:13</td>
<td>the largest we've gotten with the unsigned integers we preferred edited by Ruchaa</td>
<td></td>
<td>Complete</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0:18</td>
<td>two to the n minus one ending number in base 2 you have represent and with sign integer the two complement we have negative two to the n minus one</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0:26</td>
<td>two to the n minus one minus one because one less positive number than the negative number. So the question you might</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0:37</td>
<td>represents a very large numbers or even decimal number, rational number like pi or the natural number e. The value of this</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0:50</td>
<td>to go back and look at scientific notation base tenth</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0:56</td>
<td>things to note the number before the decimal point were call the mantissa</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1:04</td>
<td>the radix or the base that you're in so we're are talking about base 2 that would be 2 and an exponent number</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1:15</td>
<td>multiple ways to represent this number, of course, like we have point 1 and ten to the negative ten. However, to make sense you want to have stuffs in normalise form</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1:27</td>
<td>based ten that just means you have one number before the decimal point and that is also true with binary</td>
<td></td>
<td>Mark as Completed</td>
<td></td>
</tr>
</tbody>
</table>
- Deployed across 2 Computer Science courses to successfully caption selected lectures
- Work distributed across participating students. Typical example: one lecture with 11 participants
  - Each student worked on between 9 and 52 segments
  - Spent between 10 and 76 minutes over 4 days

*(hour long lecture takes around 8-10 hours to caption by an untrained person)*
Q: The captions and transcript helped me understand what the professor was saying

Q: The videos with caption (& transcript) are preferable than videos without them

Most helpful for: Learning, Attention, Efficiency, Note-taking

Somewhat helpful for: Motivation, Quiz Performance
Question: The captions and the transcript represented accurately what the professor said
Question: The ICS Caption Editor is easy to use.
Question: The placement (position) of the following elements and controls on the Caption Editor interface was appropriate?
Would you be interested in working with other students to correct captions for your class lectures using this caption editor if you receive some incentive (for example academic credit)?
ICS Videos: Usage and Experience

Employed in over 100 class sections at UH over the past 5 with many 1000s of student users total!

Freely available

Recently being employed as the technology to enable flipped classrooms (UH TIP Award 2014)
**Inverted Classroom**

Dr. Leigh Leasure – Physiological Psychology

- **Agree strongly**: A significant percentage of students strongly agreed that they prefer the combination of video and in-class discussion over a traditional class.
- **Agree**: A moderate percentage of students agreed that the review of video lectures was important to follow classroom discussions.
- **Agree slightly**: A smaller percentage of students agreed that the video was a more effective way to cover lecture material than a face-to-face lecture.

- **Disagree slightly**: A small percentage of students disagreed slightly with the combination of video and in-class discussion.
- **Disagree**: A smaller percentage of students disagreed with the review of video lectures.
- **Disagree strongly**: A minimal percentage of students strongly disagreed with the video being a more effective way to cover lecture material.

---

- I prefer the this combination of video and in-class discussion over a traditional class
- Review of video lectures was important to follow classroom discussions
- The video was a more effective way to cover lecture material than a face to face lecture
Videos & Attendance

Students who watched at least one video had higher attendance rates than those who did not watch any videos.

Video use and attendance were slightly positively correlated.
88% agreed that it is important to attend class regardless of whether or not videos were available.
Comments in surveys and focus groups were positive beyond expectations (& very satisfying)

“I can tell that I learned 80% of the class by videos”

“This is a great tool for us as a student to go over everything. I would hope that this would be available in all classes”

”Indexing enabled me to jump directly to my trouble spots”.

”I’m used to listening and then looking at the caption because it helped me tremendously to learn the new words and vocabulary”

”The search feature function have instant results, like a Google search box”
Conclusions (sort of)

Classrom Videos are an important learning companion (akin to a textbook)

- Automatic methods are effective at making video content more accessible
- CS research challenges remain
  - Smarter/Semantic Indexing and Search
  - Merging of audio, text and image information
  - HCI for a better learning experience
- Ongoing efforts to make the technology widely available
Co-conspirators

CS Faculty: Zhigang Deng, Olin Johnson, Shishir Shah, Rakesh Verma, Christoph Eick

Students: Tayfun Tuna, Varun Varghese, Mahima Joshi, Tuhin Dey (Education)
X-Students: J. Li, C. Yun, G. Bhatt, T. Tuna. A. Verma, R. Kushalnagar, Rucha Borgaonkar,

NSMIT Staff: S. Baez-Franceschi, Pradeep Krishnan, Andrea Arias

Assessment: Lecia Barker (UT Austin), Yumei Liu, Chris Hovey (NorthEastern)

Deployment, usage and assessment
  UH Computer Science, UH Geosciences, UH Biology and Biochemistry
  UH Downtown  (Richard Alo), Texas School for the Deaf  (David Coco)
ICS Videos System

www.icsvideos.uh.edu

Username: student
Password: icsstudent
Contact email: icsvideoscontact@gmail.com
or jaspal@uh.edu