Text Based Indexing to Ease Navigation in Lecture Video
Tayfun Tuna, Varun Varghese and Jaspal Subbhlok

Background
- Videos of classroom lectures have proven to be a popular and versatile learning resource.
- A major weakness of recorded lecture videos is the inability to quickly access the content of interest.
- “Indexed Captioned Searchable (ICS) Videos” framework aims to provide quick access to video content of interest by ICS:
  - Indexing: Segmented videos
  - Search: Keyword search in video
  - Captioning: Scrolling text for audio

What is Video Indexing?
- Videos are automatically divided into logical segments, each represented by a visual index snapshot.
- User can access/switch to these segments without watching whole video.

Value Of Video Indexing

<table>
<thead>
<tr>
<th>Agree strongly</th>
<th>Agree</th>
<th>Agree slightly</th>
<th>Disagree slightly</th>
</tr>
</thead>
<tbody>
<tr>
<td>The index made it easy to navigate the video.</td>
<td>The index points separated a lecture into logical segments.</td>
<td>The index points started with a new subtopic of the lecture.</td>
<td></td>
</tr>
</tbody>
</table>

Research Question
- What is the best way to do video indexing to provide conceptual segmentation which each index points will represent subtopics?

How To Do Video Indexing?
- Index points should be:
  - Meaningful: can represent subtopics
  - Not too many: scrollable
  - Not too few: broad
- Video indexing requires:
  - Identifying Transition Points (TP) where video scene changes.

Example Steps for Video Indexing
- 1. Select “desired number of IP”.
- 2. Set “min segment duration”.
- 3. Find the segment with the smallest duration.
- 4. Compare the color/text difference with left and right and merge:
  - IF (diff(current,left)) > (diff(current,right))
  - THEN merge (current,left)
  - ELSE merge (current,right)
- 5. Repeat 3-4 until:
  - smallest_segment_duration > min segment duration
  - and total number of segments == desired number of IP

Algorithm for Video Indexing

Text Based Video Indexing
- Text on the video frames is extracted using OCR technology.
- The similarity between video sections is determined by analyzing term-frequency vector of the text sections.
- Each frame is vector of frequency of the all words.

Evaluation of Text Based Indexing
- 25 diverse lecture videos were selected from computer science, biology and geology and were indexed manually to determine the ground truth.
- Average of 75 minutes per video
- Total 30+ hours of video

Conclusion
- Text based indexing algorithm provides far more accuracy than image based and random indexing algorithms, 73% vs. 54% and 44%.
- Text based indexing was successfully used to index over hundreds of videos and got positive feedback from user surveys.
- Text based indexing is integrated with Indexed Captioned Searchable (ICS) Videos framework that includes indexing, search, and captioning in video playback and has been used by dozens of courses and 1000s of students.

Challenges and Future Work
- Incremental slide progress, irrelevant text appearing in a concept, image dominated slides with little texts are found as some challenges for finding the correct index points. Instead of comparing the slide with immediate left and right, comparing it to all slides in both sides in a weighting schema (so that closer frames will have more effect) is proposed to overcome these challenges.
- Each video has its own profile (# of words per slide, duration per slide etc.). A machine learning approach to define thresholds for different profiles is expected to increase text based indexing accuracy.