Indexing and Keyword Search to Ease Navigation in Lecture Videos

Tayfun Tuna
Jaspal Subhlok, Shishir Shah, Varun Varghese, Olin Johnson

University of Houston
Department of Computer Science
- Usage of video to deliver coursework online: commonplace

- Videos at UH; - for many years - surveys
Video lectures: powerful and versatile resource

- Professor's lecture notes
- The lecture videos
- Students' notes
- Textbook reading sources

The biggest weakness of the video format: the inability to quickly access the content of interest
AGENDA

- Introduction
- ICS Video Project
- Video Indexing
- Keyword Search
- Conclusion
ICS Video Project: quick access to video content

- **Indexing:** Segmented videos
- **Search:** Keyword search in video
- **Captioning:** Scrolling text for audio
Introduction

ICS Video Project

Video Indexing

Keyword Search

Conclusion

Video Indexer

ICS Video Player

Database

OCR

Video

Index Points

SEGMENTATION

COLOR INVERSION
Example

One of the simplest types of linear convolutions is the **operation** (or averaging filter).

Conceptually, each image pixel is replaced by the average within a square "window":

This may be expressed, **at most points** (without proving it here) as the wraparound convolution of the image with another image of a square with intensity $1/M$, where $M = \#$ pixels in the square:
AGENDA

- Introduction
- ICS Video Project
- Video Indexing
- Keyword Search
- Conclusion
Video: sequence of images

- 3 indexing algorithms
  - Canonical
  - Sequential
  - Binary
Canonical Indexing Algorithm

- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
 Canonical Indexing Algorithm

- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
**Canonical Indexing Algorithm**

- **Input:** A series of frames (Video)
- **Output:** Transition Points
- **Sequentially Iterate through entire video**
- **Compare** $i$th frame with $(i+1)$th frame
Canonical Indexing Algorithm

- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
Canonical Indexing Algorithm

- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
Canonical Indexing Algorithm

- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
Canonical Indexing Algorithm

- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame
Canonical Indexing Algorithm

- Input: A series of frames (Video)
- Output: Transition Points
- Sequentially Iterate through entire video
- Compare ith frame with (i+1)th frame

Transition Point
Canonical Indexing Algorithm

- Advantage
  - Capture all slide transitions

- Disadvantages
  - Takes a lot of time, not practical to implement
  - All transition points may not be interesting
“Sampling” in Sequential and Binary

SAMPLE INTERVAL = 2
Sequential Indexing Algorithm

JUMPING INTERVAL = 4
Sequential Indexing Algorithm

JUMPING INTERVAL = 4
Sequential Indexing Algorithm

JUMPING INTERVAL = 4

A A A A A A A A B
Sequential Indexing Algorithm

JUMPING INTERVAL = 4

Transition Point
Sequential Indexing Algorithm

- **Advantage**
  - Faster

- **Disadvantages**
  - Miss some Transition Points: sampling
  - If instructor returns back
Binary Search Indexing Algorithm
Binary Search Indexing Algorithm
Binary Search Indexing Algorithm
Binary Search Indexing Algorithm
Binary Search Indexing Algorithm
Transition points filtering : Index points

- large number of transition points
- hard to manage
- filter transitions to index points
  i) max # index points : 15
  ii) minimum length of index points : 5 minutes
Frame-to-frame comparison

- Compare corresponding pixels

- Two pixels are different:
  Color Intensity difference > Intensity Threshold

- Two frames are different:
  Number of Different Pixels > Difference Threshold
Intel 2.4Ghz quad core Desktop PC
32 bit Windows Vista O/S
Introduction
ICS Video Project
Video Indexing
Keyword Search
Conclusion

Video Indexer

Database

OCR

Video

Index Points

SEGMENTATION

COLOR INVERSION

Video Indexer

Database

OCR

Video

Index Points

SEGMENTATION

COLOR INVERSION
Computation of Wrapped Convolution

Direct computation of

\[ I(i, j) = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} I_1(m, n) \cdot I_2[(i-m)N, (j-n)N] \]

Is quite simple, although very time-consuming.

Pseudo-code:

```c
int I1[N][N], I2[N][N];

I(i, j) = 0;

I(i, j) = I(i, j) + I1(m, n) * I2[(m mod N), (j-n) mod N];

while (0 <= m < N-1);
```

Just a set of nested do-loops.

If \( N \) is large (say 512 x 512) then

for each of \( N^2 \) coordinates: \( N^2 \) additions and \( N^2 \) multiplies.

For \( N = 512 \), this is 262,144 floating point operations.
Lot of OCR tools

Google
Adobe
TESSARACT OCR
Microsoft Office

Can be integrated to ICS video project?
How much accuracy on ICS images?
**Question 3**

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

<table>
<thead>
<tr>
<th>TESSERACT</th>
<th>GOOCR</th>
<th>MODI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where did the story say that there was a statue raised in Mrs. Bethune’s honor? _B-Nik is</td>
<td>Where did the story say that there was a statue raised in Mrs. Bethune’s honor? ;_n ton Dc , m.<em>am.</em> F_or. ___a Mayetvil South0rolina</td>
<td>Where did the story say that there was a statue raised in Mrs. Bethune’s honor? Washinton, D.C.Miami, Florida Mayesville, South Carolina</td>
</tr>
</tbody>
</table>
Challenges of OCR Tools

- OCR tools: black and white images
- ICS VIDEO Images: complex shapes and different colors
- OCR tools: fonts 10-72 pixels
- ICS VIDEO Images: smaller fonts like captions in a graph
Examples from ICS images

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.
Segmentation of Text Region Steps

1. Input Image
2. Binarization (SIS)
3. Dilation Effect (x8)
4. Edge Detection (Sobel)
5. Blob Extraction (Labeling)
6. Resize to 1.5x (Interpolation)
Segmentation of Text Region Steps

a) Original image
b) Binarization
c) Dilation effect
d) Edge detection
e) Blob extraction
f) Resizing
### Inversions of Colors

<table>
<thead>
<tr>
<th>Original Image</th>
<th>R / G / B</th>
</tr>
</thead>
</table>

| Source code | Interpreter | Results |

Input (source code): program in a computer programming language. The code is translated to object code and immediately executed. Some programming languages are interpreted rather than compiled.

---

| 255-R / 255-G / 255-B |

| Source code | Interpreter | Results |

Input (source code): program in a computer programming language. The code is translated to object code and immediately executed. Some programming languages are interpreted rather than compiled.
# Inversions of Colors

<table>
<thead>
<tr>
<th>Original Image</th>
<th>R / G / B</th>
<th>Inversion1</th>
<th>255-R / G / B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Original Image" /></td>
<td>R / G / B</td>
<td><img src="image2.jpg" alt="Inversion1" /></td>
<td>255-R / G / B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inversion2</th>
<th>R/ 255-G / B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3.jpg" alt="Inversion2" /></td>
<td>R/ 255-G / B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inversion3</th>
<th>R/ G / 255-B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.jpg" alt="Inversion3" /></td>
<td>R/ G / 255-B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inversion4</th>
<th>255-R/ 255-G / B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image5.jpg" alt="Inversion4" /></td>
<td>255-R/ 255-G / B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inversion5</th>
<th>R/ 255-G / 255-B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image6.jpg" alt="Inversion5" /></td>
<td>R/ 255-G / 255-B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inversion6</th>
<th>255-R/ G / 255-B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7.jpg" alt="Inversion6" /></td>
<td>255-R/ G / 255-B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inversion7</th>
<th>255-R / 255-G / 255-B</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image8.jpg" alt="Inversion7" /></td>
<td>255-R / 255-G / 255-B</td>
</tr>
</tbody>
</table>
OCR Tool Experiments with Transformations

- 20 different videos
- 1387 different images (transition points)
- 20007 unique words
- 27201 total words

Search Accuracy = \[
\frac{\text{# Detected Unique Words}}{\text{# Total of Unique Words}}
\]
OCR Tool and Image Transformation Execution Times

Segmentation + Inversion

Inversion

Segmentation

OCR Time

Inversion: 8 OCR process
Segmentation: 1-20 OCR process

Intel 2.4Ghz quad core Desktop PC
32 bit Windows Vista O/S
AGENDA

- Introduction
- ICS Video Project
- Video Indexing
- Keyword Search
- Conclusion
Conclusion

- Indexing and OCR Tool: computationally efficient:
  1-hour video process time < 15 minutes
- Indexing algorithm near perfect in catching transition points -- 96% accuracy
- OCR tool detection and accuracy enhanced with image transformations -- from ~91% to ~97%
- The ICS video framework used by dozens of courses in the past few semesters, at two UH campuses
- Assessments for ICS Framework found favorable:
  Indexing and Search features very helpful and easy to use but large scale validation is ongoing.
THANK YOU

http://icsvideos.cs.uh.edu/
Transition points Detection Performance

JI : 64 and SI: 2

Transition Points

Binary
Sequential
Canonical
Execution Time Performance

Intel 2.4Ghz quad core Desktop PC
32 bit Windows Vista O/S

JI : 64 and SI: 2
## False Positives

<table>
<thead>
<tr>
<th></th>
<th>Modi</th>
<th>Gocr</th>
<th>Tesseract</th>
<th>Modi-Gocr-Tesseract</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCR</td>
<td>19256</td>
<td>10363</td>
<td>13610</td>
<td>45448</td>
</tr>
<tr>
<td>Seg+OCR</td>
<td>26472</td>
<td>27005</td>
<td>32557</td>
<td>64590</td>
</tr>
<tr>
<td>Inv+OCR</td>
<td>76620</td>
<td>52768</td>
<td>80837</td>
<td>161192</td>
</tr>
<tr>
<td>Seg+Inv+OCR</td>
<td>88563</td>
<td>67322</td>
<td>105347</td>
<td>200430</td>
</tr>
</tbody>
</table>
## # of Detection Increase

<table>
<thead>
<tr>
<th>Method</th>
<th>FirstUndetectThenDetectTotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modi vs IE_Seg+Modi</td>
<td>613</td>
</tr>
<tr>
<td>Modi vs IE_Inv+Modi</td>
<td>1057</td>
</tr>
<tr>
<td>Modi vs IE+Modi</td>
<td>1131</td>
</tr>
<tr>
<td>Gocr vs IE_Seg+Gocr</td>
<td>1926</td>
</tr>
<tr>
<td>Gocr vs IE_Inv+Gocr</td>
<td>2347</td>
</tr>
<tr>
<td>Gocr vs IE+Gocr</td>
<td>3174</td>
</tr>
<tr>
<td>Teseract vs IE_Seg+Tesseract</td>
<td>1521</td>
</tr>
<tr>
<td>Teseract vs IE_Inv+Tesseract</td>
<td>1693</td>
</tr>
<tr>
<td>Teseract vs IE+Tesseract</td>
<td>2355</td>
</tr>
<tr>
<td>ModiGocrTesseract vs IE_Seg+ModiGocrTesseract</td>
<td>214</td>
</tr>
<tr>
<td>ModiGocrTesseract vs IE_Inv+ModiGocrTesseract</td>
<td>523</td>
</tr>
<tr>
<td>ModiGocrTesseract vs IE+ModiGocrTesseract</td>
<td>513</td>
</tr>
</tbody>
</table>
Future Work

- More surveys of ICS framework
- Logging the usage of features
- Enable searching inside multiple videos
- Using multithreading on indexing and OCR tools
- Word completion with predefined dictionaries
- “Keyword Search” to “Semantic Search”