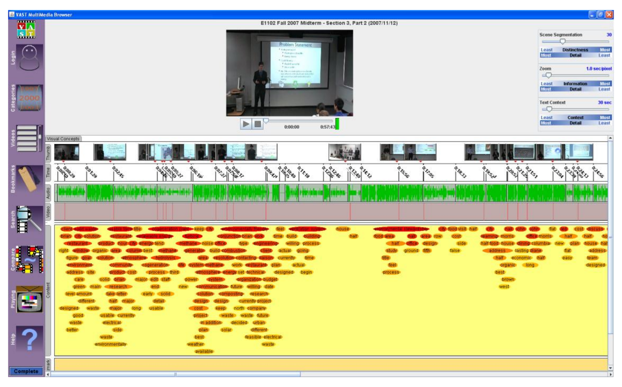
Semantic Keyword Extraction via Adaptive Text Binarization of Unstructured Unsourced Video

Michele Merler and John R. Kender
Computer Science Department
Columbia University
{mmerler,jrk}@cs.columbia.edu

Introduction

- Videos of presentations are employed in a large variety of systems for different purposes
  - Distance or E-learning
  - Automatic generation of conference proceedings
  - Student presentations



- Challenges

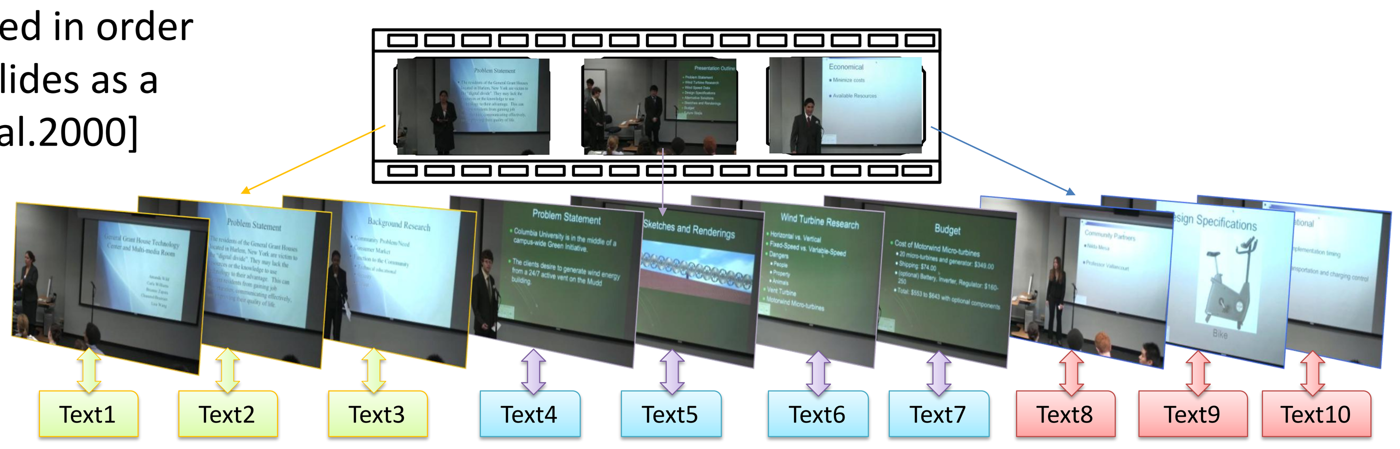
Many videos are already archived	Low quality	Lack of Structure
Lack of additional sources of information (e.g. electronic copies of slides)	Unconstrained camera movements; Slides Clipped; Compression	Not recorded by professional cameramen; Shots cannot be used as clue; Not edited

- Result: Fully automatic method for summarizing and indexing unstructured presentation videos based on text extracted from the projected slides
- Integration into summarization and presentation tools such as the VAST MultiMedia Browser[1] (see image above)

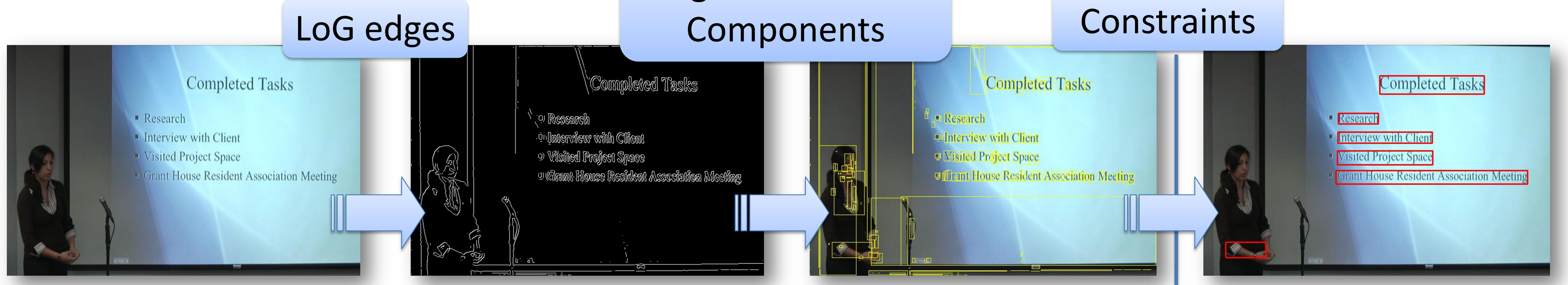
Proposed Approach

1. Segment video into semantically distinct shots based on slides

- Studies have been conducted in order to assess the reliability of slides as a summarization tool [He et al.2000]
- No electronic copies of the slides NEW!
- Changes in text used to assess slide changes



2. Slides Text Detection



Geometric Constraints

$$\frac{F_{Area}}{1000} \le R_{Area} \le \frac{F_{Area}}{10}$$

$$2 \le R_{width} \le \frac{F_{width}}{3}$$

$$6 \le R_{height} \le \frac{F_{height}}{5}$$

Alignment Regions Merge Constraints

$$merge(R_i, R_j) \Leftrightarrow \begin{cases} R_i \cap R_j > 0 \\ |x(R_i) - x(R_j)| \le 10 \end{cases}$$

Edge Density Constraint

$$E_{density} \ge 0.2$$

Empirically validated thresholds applied to prune non-text regions

F – Frame
R – Candidate Text Region

3. Slides Text Recognition

- Double Text Regions Size with Bilinear Interpolation
- LAO Binarization
- Tesseract[2] OCR Engine
- Training with 15 character sets
  - Most popular fonts used in presentation slides
  - Text reflecting English letters frequencies[3]
- Height 30pt

NEW!

Local Adaptive Otsu (LAO) Binarization

NEW!

Optimize for threshold *T* maximizing between-class variance in **sliding window**

Optimal version of Sauvola's algorithm

$$T(x,y) = \mu(x,y,W)\left[1 + k\left(\frac{\sigma(x,y,W)}{R} - 1\right)\right]$$

[Sauvola et al. 00]

Dependency from *k* is removed!

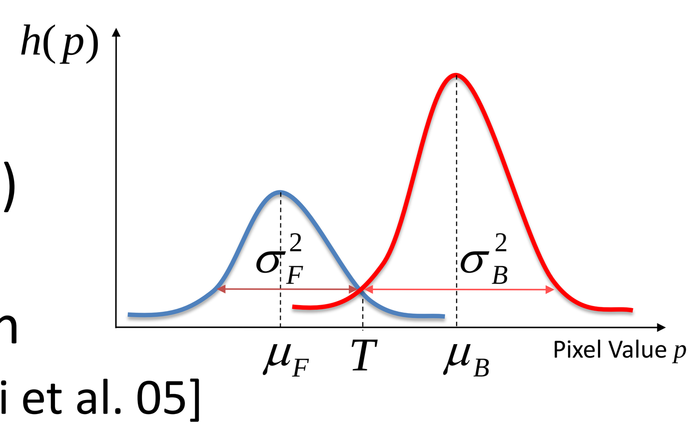
Localized version of Otsu's algorithm [Otsu 79]

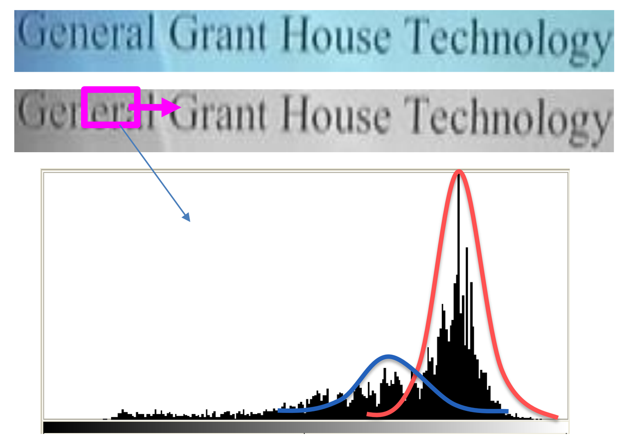
$$\sigma^2_{between}(T) = n_B(T)n_F(T)(\mu_B(T) - \mu_F(T))^2$$

$$\sigma^2_{within}(T) = n_B(T)\sigma^2_B(T) + n_F(T)\sigma^2_F(T)$$

[Otsu 79]

- Assumption: bimodal distribution (foreground/background)
- Fast implementation with Integral Histogram [Porikli et al. 05]





Results

8 presentation videos, 1hr 45 mins, ~13 Slides each

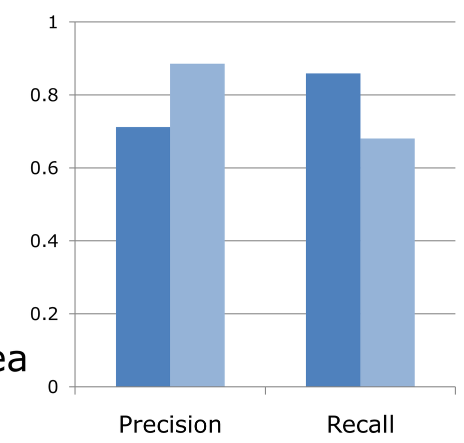
Slides Text Detection - 500 Frames

400 with text
100 no text

$$Precision = \frac{TA_{GT} \cap TA_E}{TA_E}$$

$$Recall = \frac{TA_{GT} \cap TA_E}{TA_{GT}}$$

TA_{GT} → Ground Truth Text Regions Area
TA_E → Extracted Text Regions Area



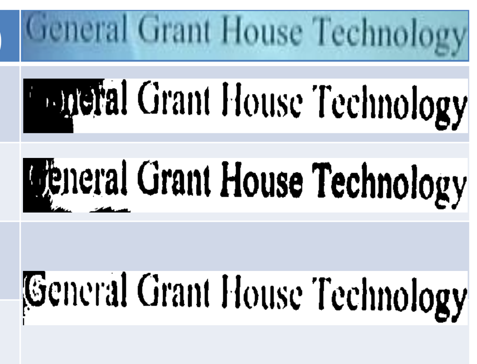
- SIMPLE – use all candidate text regions
- REFINED – prune candidate text regions based on OCR output-

$Precision_{SIMPLE}$	$Recall_{SIMPLE}$
0.71213	0.85914

$Precision_{REFINED}$	$Recall_{REFINED}$
0.88584	0.68046

Binarization - 54 Detected Regions, 2177154 annotated pixels

Algorithm	Precision	Recall	F1*	t(sec)
Otsu	0.8611	0.8555	0.8583	0.539
Sauvola (k = 0.5)	0.9003	0.8759	0.8879	0.626
LAO	0.8831	0.9278	0.9049	2.126
LAO + Int. Hist.	**0.8831**	**0.9278**	**0.9049**	**1.29**



Text vs. Non-Text
Foreground vs. Background

$$F1^* = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Slides Text Recognition - 2276 words, 13804 characters

$$Precision = \frac{N_{cor(c/w)}}{N_{gt(c/w)}} \quad Recall = \frac{N_{cor(c/w)}}{N_{r(c/w)}} \quad N_{corc} = N_{gtc} - ED(s_{gt}, s_r)$$

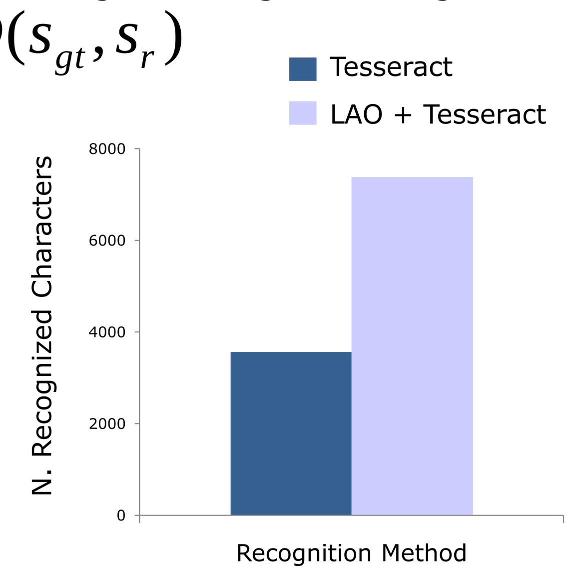
Edit Distance between Ground Truth String and Recognized String

Character Recognition

N. Ground Truth Characters	N. Recognized Characters	Precision	Recall
13804	7376	0.5343	0.7446

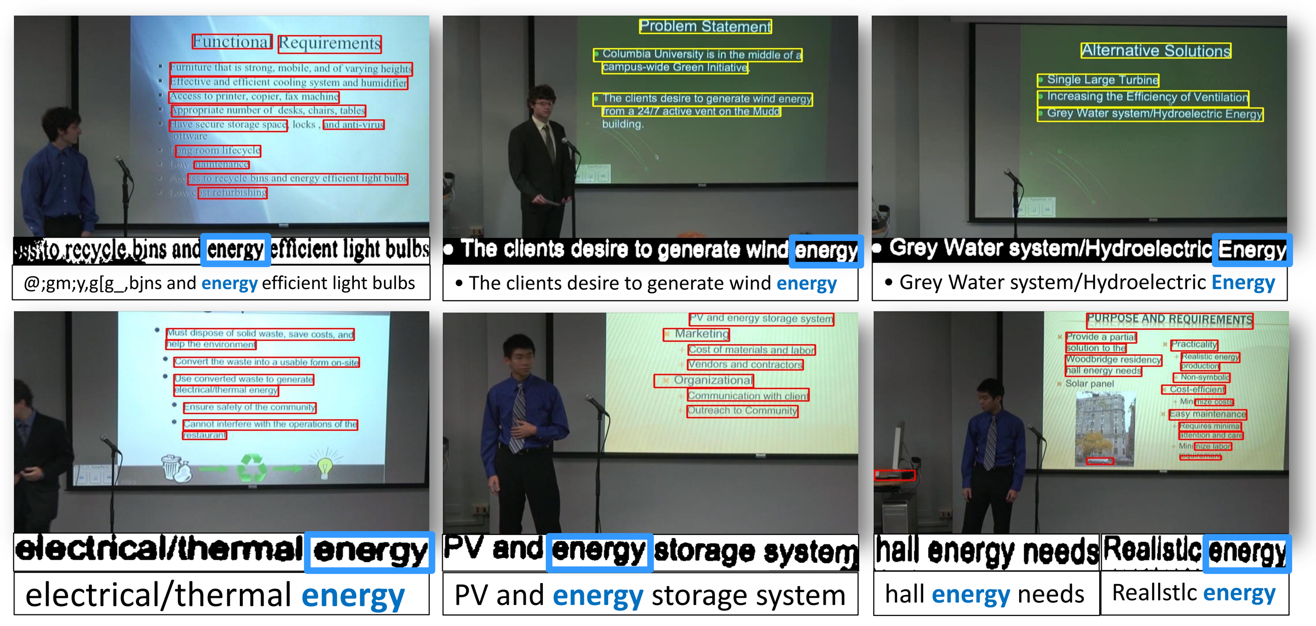
Word Recognition

N. Ground Truth Words	N. Recognized Words	Precision	Recall
2276	1126	0.4947	0.6651



Example of indexing function:

the word Energy is recognized in slides across 4 different presentations



1. www.aquaphoenix.com/research/vastmm 2. http://code.google.com/p/tesseract-ocr 3. http://en.wikipedia.org/wiki/Letter_frequencies