

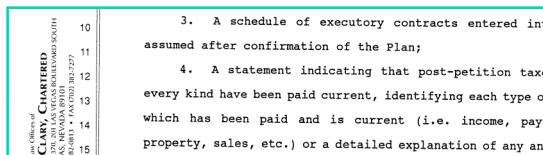
# A Discriminative Semi-Markov Model for Robust Scene Text Recognition

Jerod Weinman



Computer Vision Laboratory  
Department of Computer Science  
University of Massachusetts Amherst

## Page Readers

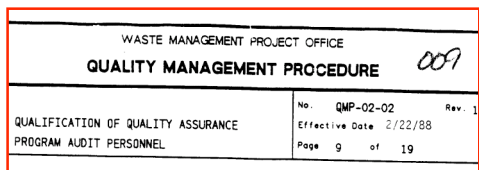


Legal

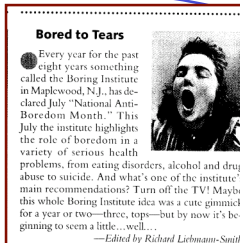
99.7%

DOE

98.4%



## Annual Rpt. Magazine



97.4%

96.6%

[Rice et al. 96]



traditional Asi  
Hcalipg Arts

Traditional Asian  
Healing Arts

## Reading Signs



Detection



traditional Asi  
Hcalipg Arts



Recognition

## What Challenges?

- Small sample (< 50 chars)

31 BOLTWOOD WALK

Via Via

- Perspective distortion

HASTINGS

DELANO'S  
Since 1977

## What challenges?

- Uneven lighting



- Unusual fonts



## What challenges?

- Word segmentation



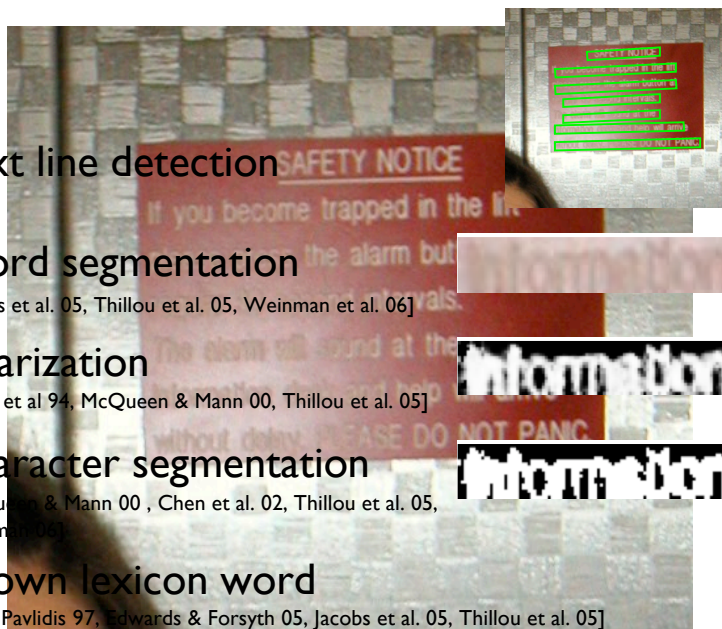
## What challenges?

- Character segmentation



## What Assumptions?

- Text line detection
- Word segmentation  
[Jacobs et al. 05, Thillou et al. 05, Weinman et al. 06]
- Binarization  
[Ohya et al 94, McQueen & Mann 00, Thillou et al. 05]
- Character segmentation  
[McQueen & Mann 00, Chen et al. 02, Thillou et al. 05, Weinman 06]
- Known lexicon word  
[Shi & Pavlidis 97, Edwards & Forsyth 05, Jacobs et al. 05, Thillou et al. 05]

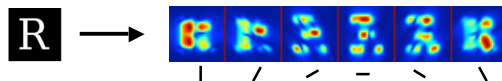


## Overview

- Our approach
  - No binarization
  - No prior word or character segmentation
  - Lexical bias, but not forced
- Recognition information
- Finding an interpretation
- Semi-Markov model
- Experiments
- End-to-End demonstration

## Recognition Information

- Appearance



- Bigrams

$$P(\text{TH} | \text{English}) = \frac{39}{1000} \quad P(\text{QU} | \text{English}) = \frac{1.4}{1000}$$

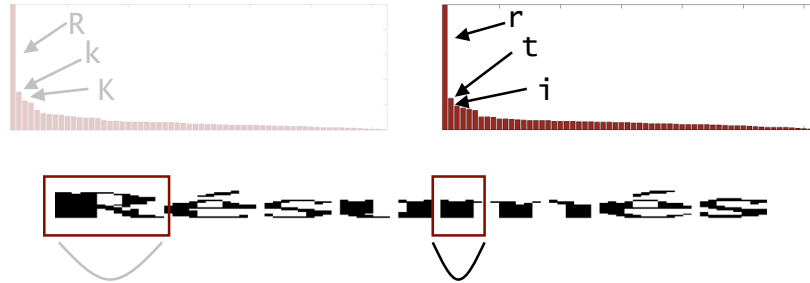
$$P(\text{IN} | \text{English}) = \frac{21}{1000} \quad P(\text{QA} | \text{English}) = \frac{.0001}{1000}$$

- Lexicon

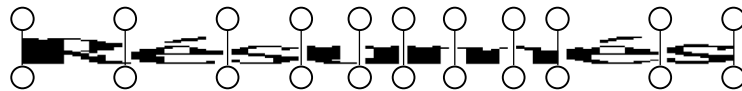


{ a  
Aaberg  
Aachen  
⋮  
zymurgy  
Zyuganov  
Zzz  
⋮

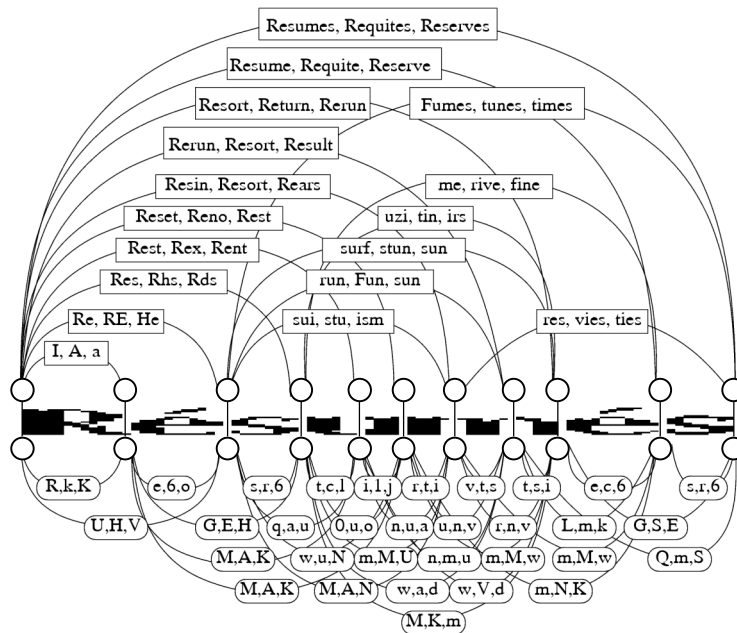
## Interpretation Scoring



## Interpretation Graph



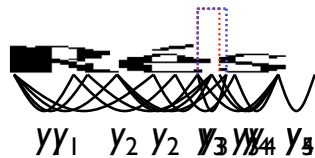
## Interpretation Graph



## Semi-Markov Model

$$p(\mathbf{y}|\mathbf{x};\bar{\theta}) = \frac{1}{Z(\mathbf{x})} \exp\{U(\mathbf{y}, \mathbf{x}; \bar{\theta})\}$$

[Lafferty et al. 01]  
[Sarawagi & Cohen 05]



- Appearance

$$U^A(y_i, \mathbf{x})$$

- Bigrams

$$U^B(y_i, y_{i+1})$$

- Lexicon

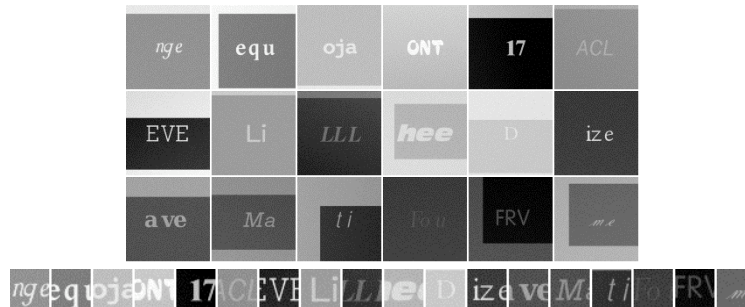
$$U^L$$

- Parse

$$U^P(y_i, y_{i+1}, \mathbf{x})$$

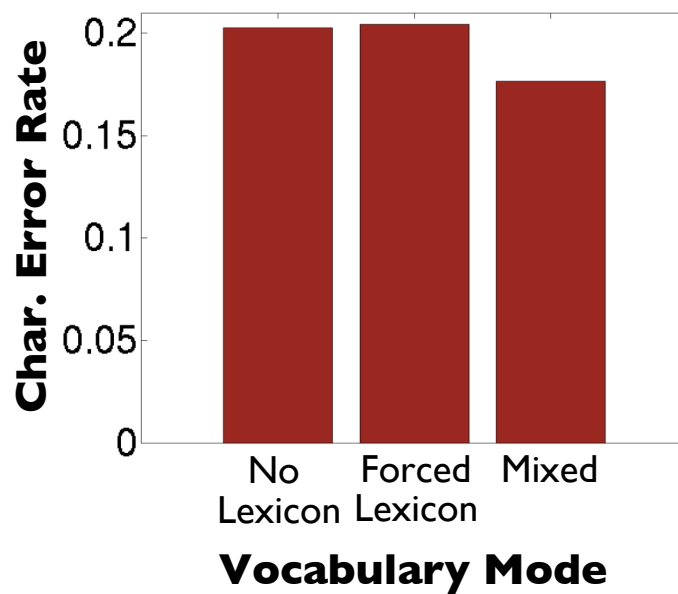
## Experimental Data

- Synthetic characters (934 fonts)



- English text (82 books, Project Gutenberg)
- Lexicon (50th freq. percentile, SCOWL)
- 85 sign images (1100 char., 12 pixel x-height)

## Results



## Recognition Comparison v. OmniPage

		uucL,ass
		Free ehe in
		LIBRFIRY
		A),1HERbT
		TAVF
		Tradiliorpal Asi'? HealiIN Arts

## Results

System	Char Err.
OmniPage	23.5%
OmniPage + Binarized	16.6%
Semi-Markov	15.0%

## Low Resolution

		Resumes
		Resumes
		WM/ I WS
		krauts
		11.-4her

† tttA wms

## Low Resolution

		Je eryAmherst
		JefferyAmherst
		JONryAmhent
		GkilyaryAmhent
		.WlervAMMnt

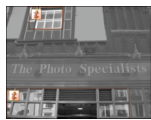
\* Lit wryA An memo

† An wrrA Am wme

## End-to-End Demonstration

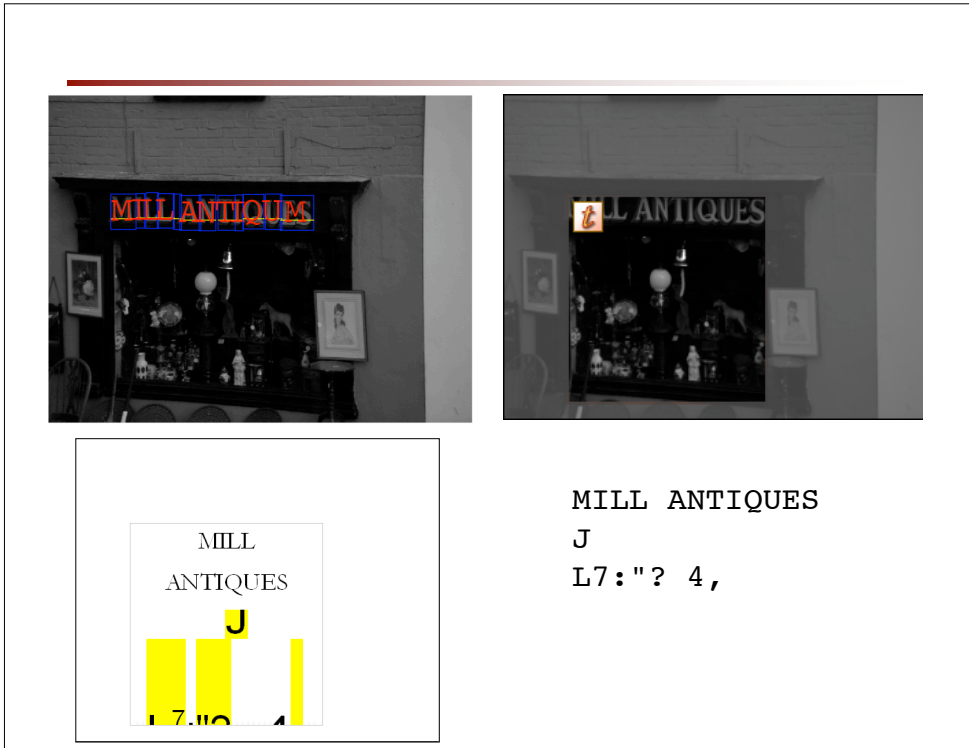
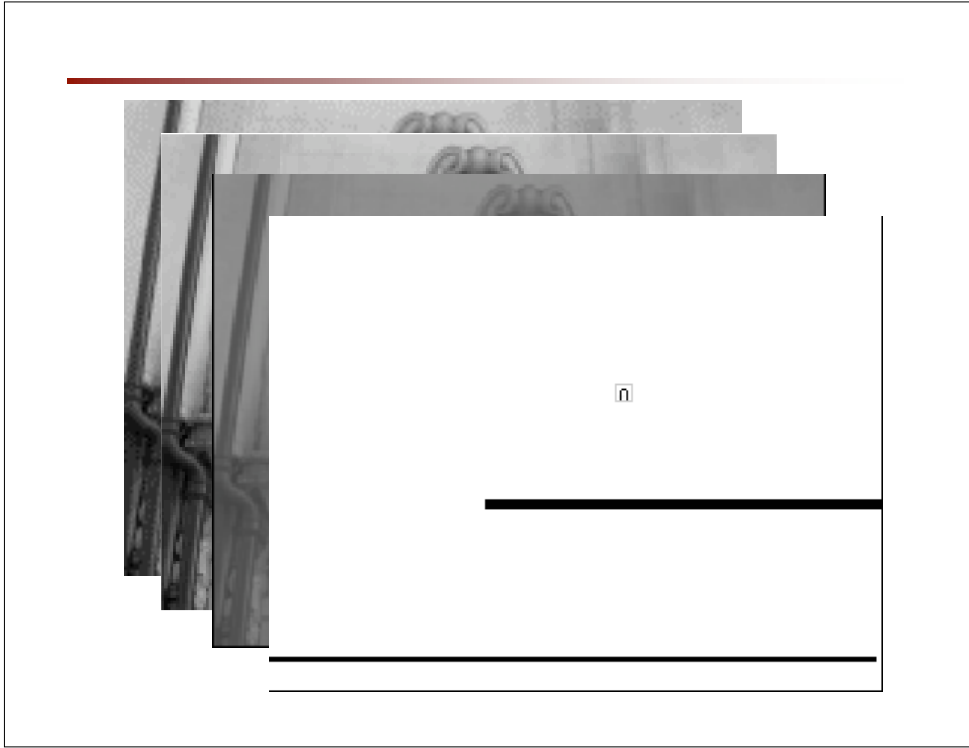


## OmniPage



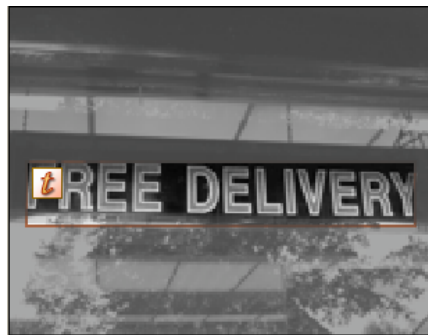
*i*   *taw*  
*aim*

1/1/   噠噠   的T  
ti/Tir-TWITE.A





FIREE DELIVEp



Ma OS

## Summary

- No binarization
- No prior word segmentation
- Joint character segmentation and recognition
- Lexical bias, but not forced

# Thank You

## **Support**

NSF IIS-0326249

NSA

NSF IIS-0100851

CIA

## **Collaborators**

Allen Hanson

Piyanuch Silapachote

Erik Learned-Miller

Marwan Mattar

Andrew McCallum

Richard Weiss