

## Fast Lexicon-Based Scene Text Recognition with Sparse Belief Propagation

Jerod Weinman  
Erik Learned-Miller  
Allen Hanson



Computer Vision Laboratory  
Department of Computer Science  
University of Massachusetts Amherst

## Camera-Based Text Recognition



### Overview

- Information sources / Features
- Model structure
- Inference
- Sparse belief propagation
- Experimental results



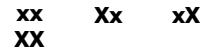
### Recognition Information

- Appearance



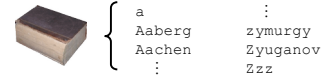
- Language

- Bigrams
- Letter case

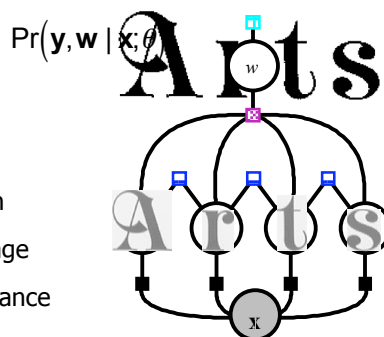


- Lexicon

- Word list
- Bias



### Recognition Model



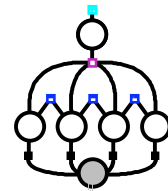
- Bias
- Lexicon
- Language
- Appearance

### Recognition Model

$$\Pr(\mathbf{y}, \mathbf{w} | \mathbf{x}; \theta) \propto \prod_c f_c$$

#### Factors / Compatibilities

- Bias  $f(w; \theta)$
- Lexicon  $f(\mathbf{y}, \mathbf{w})$
- Language  $f(y_i, y_{i+1}; \theta)$
- Appearance  $f(y_i, \mathbf{x}; \theta)$



### Lexicon Factor & Inference

$$f(\mathbf{y}, w) = \begin{cases} w = 1 \wedge \mathbf{y} \notin L & 0 \\ \text{otherwise} & 1 \end{cases}$$

- “Bottom-Up”  
Product of Messages  
 $b(y_2) \approx p(y_2 | \mathbf{x}; \theta)$
- “Top-Down”  
Sum-Product  
14 million strings    16K words    23 words

$$m'(y_2) = \sum_{\mathbf{y} \vdash y_2, w} f(\mathbf{y}, w) \cdot m(\mathbf{y}, w)$$

### Sparse Belief Propagation

[Pal, Sutton, and McCallum 2006]

$$m'(y_1) = \sum_{\mathbf{y} \vdash y_1, w} f(\mathbf{y}, w) \cdot m(\mathbf{y}, w)$$

$\text{KL}(b'(y_2) || b(y_2)) \leq \epsilon$

### Shortening the Lexicon

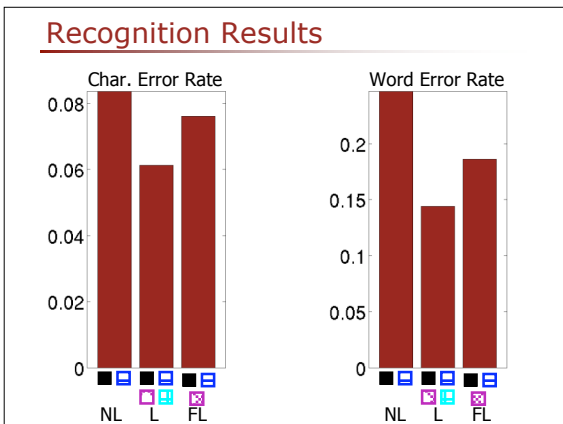
Character

### Sparsity Impact

### Recognition Examples

### Recognition Examples

	USED BOOKS	HOOK-UPS	SI BOLTWOOD	RELANO3
No Lexicon	USED BOOKI	HTOR UP5	31 BOLTWOOD	RELAmo3
Lexicon	USED BOOKS	HOOK UPS	31 BOLTWOOD	RELAmo3
Forced Lexicon	USED BOOKS	HOOK UPS	SI BENTWOOD	DELANOS



- ### In Conclusion
- Integrating lexical decision is beneficial
  - Sparse BP accelerates inference
  - Ongoing work
    - Integrating more information
    - Joint detection and recognition

### Thank You

**Support**



NSF



NSA



CIA

**Collaborators**

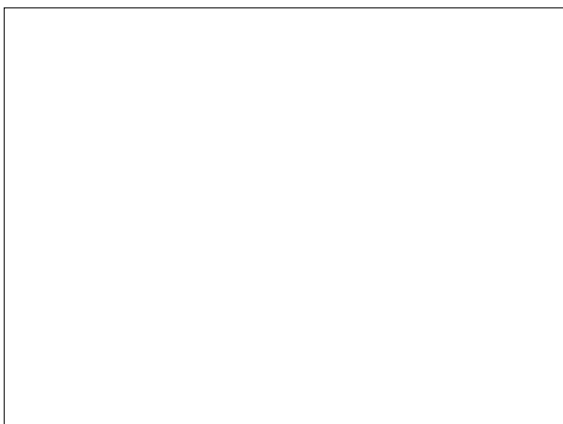
Andrew McCallum  
Piyanuch Silapachote  
Marwan Mattar

Chris Pal  
Charles Sutton

### Recognition Examples

The 'Correct' category shows a grid of various text samples including: FIBER ART CENTER, HASTINGS GREETING CARDS, FIRE Resumes, Jeffery Amherst College Store, 24 Hour Banking, Loose Goose, AMHERST MARTIAL ARTS, RAVEN, Clark House, Traditional Asian Healing Arts, JUICE BAR, Restaurant, SCUP BOWL DOUGLAS CHINA INN, INSURANCE, COFFEE HOUSE, BANANARAMA SAINT BRIGIDS, CONTEMPORARY CRAFTS, FIRE, UNITED STATES POSTAL SERVICE, THE EVERGREENS, Upstairs, TOWN HALL, PHONE CARDS.

The 'Hard' category shows a single sample: PRODUCTS FOR LIFE IN THE FAST LANE, MYSTERY TRAIN RECORDS.



- ### Training Details
- Appearance
    - 200/800 Training and validation fonts
    - Laplace Prior
  - Language
    - Corpus of 82 Project Gutenberg books
    - 49 million characters
  - Lexicon
    - SCOWL 70th percentile
    - 187,000 words