

Unified Detection and Recognition for Reading Text in Scene Images

Jerod Weinman



Computer Vision Laboratory
Department of Computer Science
University of Massachusetts Amherst



Herb Simon and Allen Newell, c. 1958



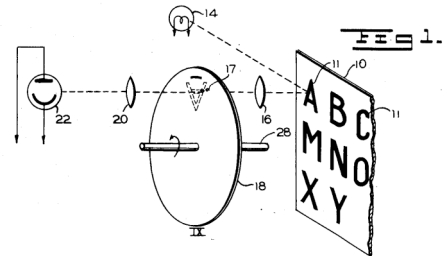
Dec. 22, 1953

D. H. SHEPARD
APPARATUS FOR READING

2,663,758

Filed March 1, 1951

6 Sheets-Sheet 1



IBM's Deep Blue



Seven-Year-Old

Rice et al., *Optical Character Recognition*, Kluwer, 1999.

Page Readers

Legal	Business	DOE
99.7%	99.4%	98.4%
Annual Rpt. Magazine		
97.4%	96.6%	

Sign and Text Detection

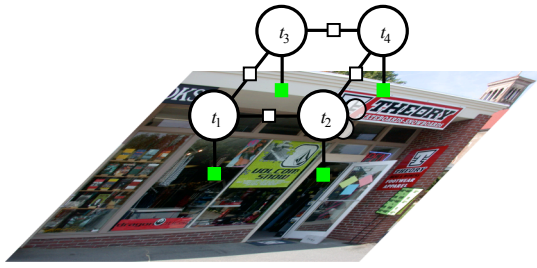


Context

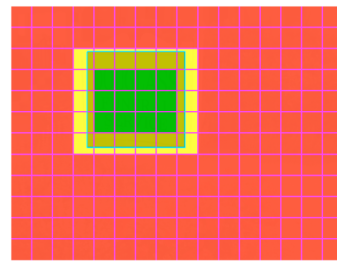


Contextual Model

$\Pr(\mathbf{t} | \text{Image}, \text{Experience})$

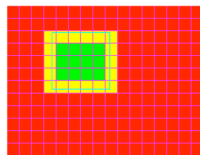


Mixed Labels



Partially Labeled Data

$$p(\mathbf{t}_{\text{Sign}}, \mathbf{t}_{\text{Bg}}, \mathbf{t}_{\text{Mix}} | \mathbf{x}; \theta)$$



Marginal Likelihood:

$$p(\mathbf{t}_{\text{Sign}}, \mathbf{t}_{\text{Bg}} | \mathbf{x}; \theta) = \frac{p(\mathbf{t}_{\text{Sign}}, \mathbf{t}_{\text{Bg}}, \mathbf{t}_{\text{Mix}} | \mathbf{x}; \theta)}{p(\mathbf{t}_{\text{Mix}} | \mathbf{t}_{\text{Sign}}, \mathbf{t}_{\text{Bg}}, \mathbf{x}; \theta)}$$

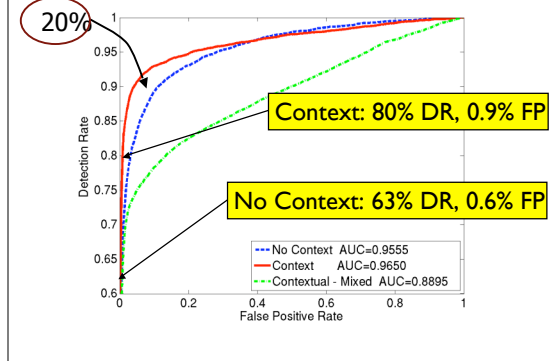
Detection Examples



Detection Examples



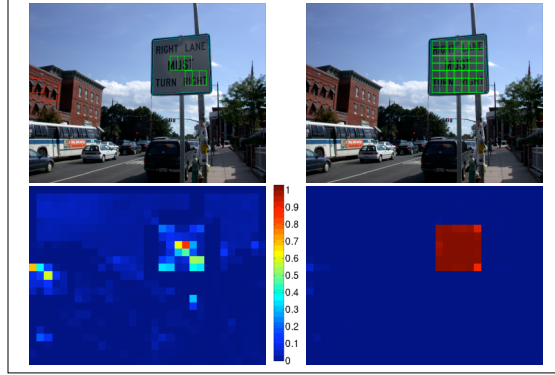
Detection Comparison



Detection Results

Sign-Level	Local	Contextual
Detection Rate	86±1%	85±2%
Avg. Coverage	74 ±2%	88 ±1%
Median Coverage	81%	100%
False Pos Area/Image	0.97 ±0.01%	1.0 ±0.2%

Detection Comparison

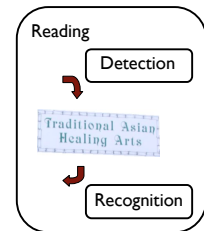


Contributions

- Learned Layout Analysis
- Simultaneous Multi-Scale Detection
- Training with Partially Labeled Data

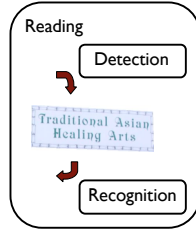
Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition



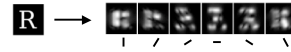
Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition



Recognition Information

- Appearance

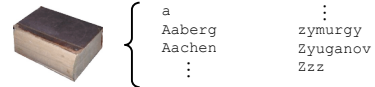


- Local Language

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

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

- Lexicon



Recognition Information

Ξαλλαλλαννα



Mississippi

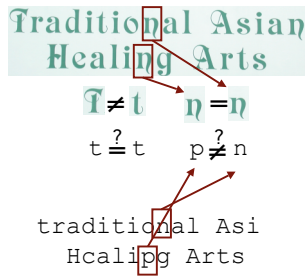
Recognition Information

- Similarity



Recognition Information

- Similarity



Learning Similarity

$$\eta \neq \eta \quad T \neq t$$

$$p \neq n \quad t = t$$

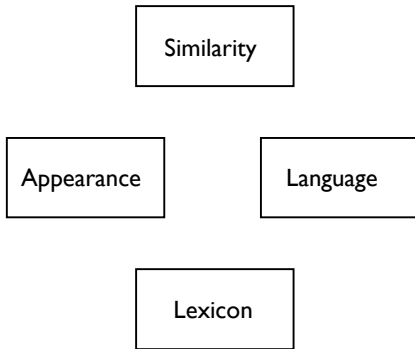
- Same



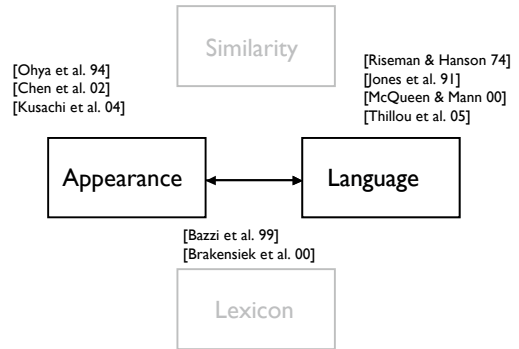
- Different



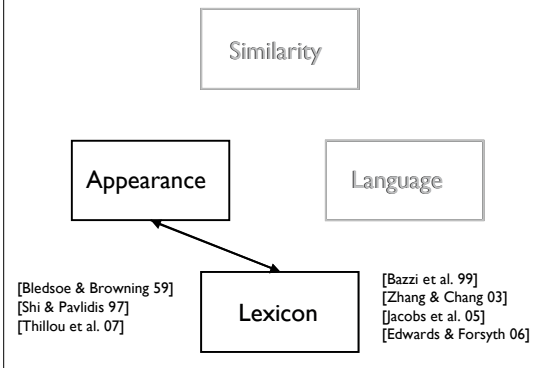
All the Information



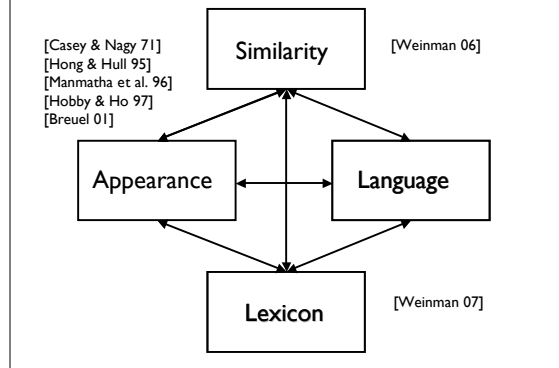
All the Information



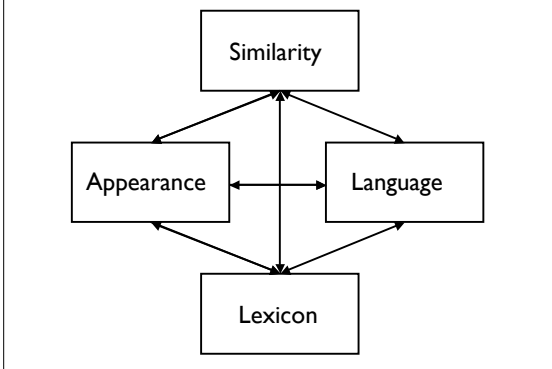
All the Information



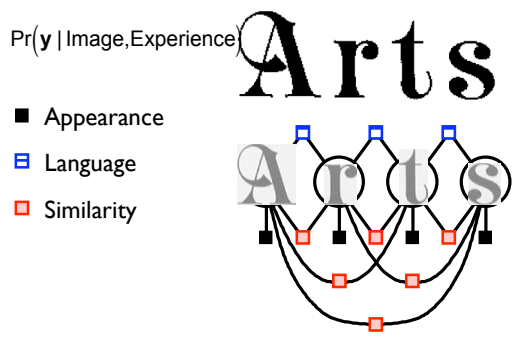
All the Information



All the Information



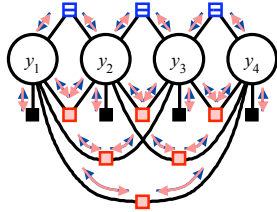
Recognition Model



Recognition Model

$$\Pr(y | \text{Image, Experience})$$

- Appearance
- Language
- Similarity



Recognition Examples

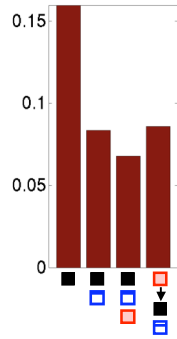


Correct

Hard

Recognition Results

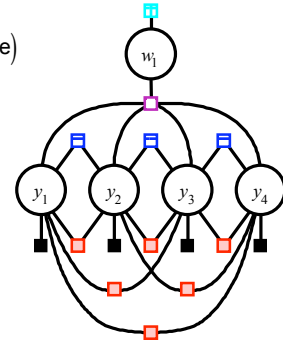
Char. Error Rate



Recognition Model

$$\Pr(y | \text{Image, Experience})$$

- Appearance
- Language
- Similarity
- Lexicon
- Bias



Shortening the Lexicon

AARA	Dreg	KEAN	Pica	WING	hoke	thaw
ADDS	Dyed	KOLK	Poke	Wean	idos	tobe
AMAL	ELNA	Kah	Pym	YENS	inss	twig
ANYS	ERLS	Kind	RAMS	WES	jeep	urpa
AVER	Edan	Kora	RICE	ZIMA	join	vile
Acts	Elia	LECH	ROOD	Zerk	keas	vows
A11a	Eurs	LIDS	Rags	afro	k1ln	webb
Anew	FANE	LOXS	Redo	alds	lads	weys
Arks	FLOR	LYND	Roeg	arch	leak	wkly
Arts	FRIS	Leus	Rude	assn	Tivy	wrap
BABS	Figs	Lirs	SEER	bark	lous	yeti
BEVY	Fobs	Lurs	SHIR	bent	mara	yore
BLOK	GARD	MALT	SOHS	boat	meme	zinc
BUDS	CEUH	MLK	SRTA	brod	mope	zuni
Badr	GRIZ	MOLT	Sair	cati	myke	
Bice	Gala	Mams	Seda	choc	nine	
B1vd	Clia	Mega	S1xs	coos	numb	
Bump	Grac	Mood	Soxs	cubs	oort	
CANA	HAZE	Mutt	Suns	demo	otti	
CION	HIGH	NIKO	TACS	ding	Peon	
COIT	HYMN	NIJRS	TICA	duck	pipe	
CUPV	Hats	Nils	TOLD	ease	puce	
Cary	Hood	Null	Tabb	emit	race	
Clap	Hutt	OORT	Tavi	espy	reus	
Cons	IRAS	OTTE	ToFF	Febs	road	
DALS	Ibid	Omar	Trev	Fitz	sabu	
DELE	Ivah	Orva	URSI	fyns	scop	
DOME	JAVA	PCEP	Urea	gaye	stls	
DRYE	JUNE	PIER	VOLA	goof	smog	
Deck	Jebs	PRDP	Vern	gulp	sued	
Diba	KABS	Paco	WEFT	herb	synd	

Character

Sparsity Impact

- Median number of characters: 4
- Median percent of words: 0.07%
- Median number of words: 16 (of 35,000)

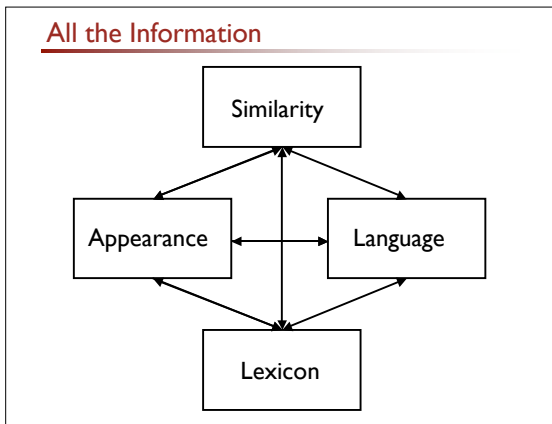
Recognition Examples

	USED BOOKS	HOOK-UPS	RELAMO'S	31 BOLTWOOD
No Lexicon	USED BOOKI	HTOR UP5	RELAMO 3	31 BOLTWOOD
Lexicon	USED BOOKS	HOOK UPS	RELAMO 3	31 BOLTWOOD
Forced Lexicon	USED BOOKS	HOOK UPS	DELANO S	SI BENTWOOD

Recognition Results



All the Information

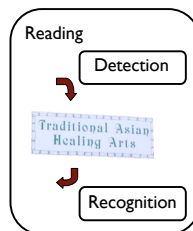


Contributions

- Unified model for recognition
- Learned similarity function for consistency
- Principled, fast lexicon integration

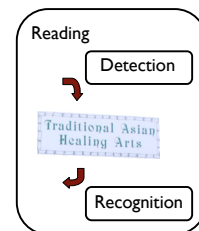
Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition



Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition



Word Segmentation

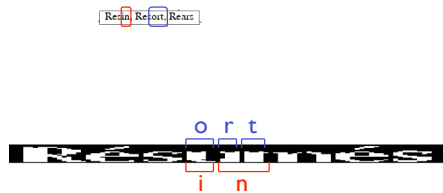
FIRE
CRAFTS
salon

Jeffery Amherst
MARTIAL ARTS
UNITED STATES

Character Segmentation



Interpretation Graph



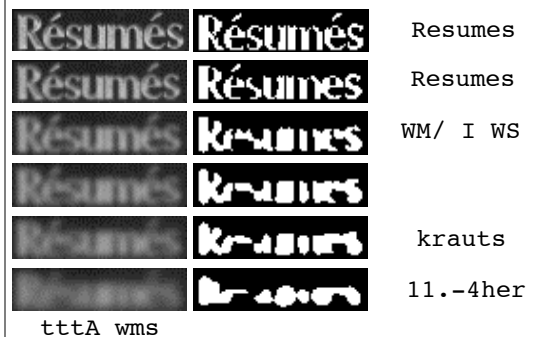
Recognition Comparison



Results

	Char Err.
OmniPage	23.5%
OmniPage + Binarized	16.6%
Ours	15.0%

Low Resolution



Low Resolution

Jeffery Amherst	Jeffery Amherst	Je eryAmherst
Jeffery Amherst	Jeffery Amherst	JefferyAmherst
Jeffery Amherst	Jeffery Amherst	JONryAmhent
Jeffery Amherst	Jeffery Amherst	GkiliaryAmhent
Jeffery Amherst	Jeffery Amherst	.WlervAmMnt
Jeffery Amherst	Jeffery Amherst	Jo 116 fp
		Ak 11_Se-

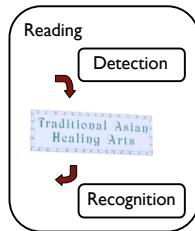
* Lit wryA An memo
† An wrrA Am wme

Contributions

- Unified word segmentation and recognition
- Hybrid open and closed vocabulary modes
- Recognition bias prevention

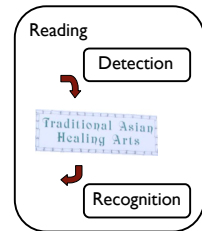
Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition



Overview

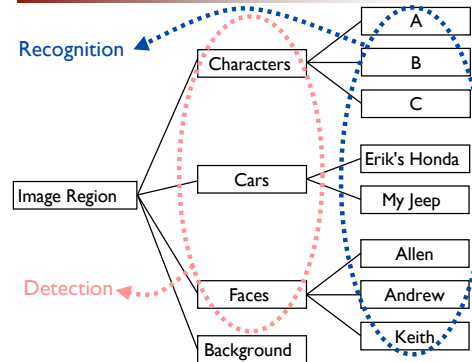
- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition



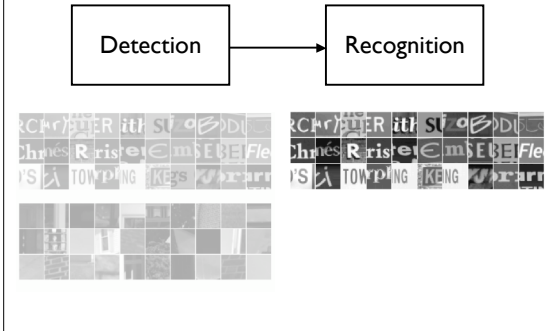
Reading Problem



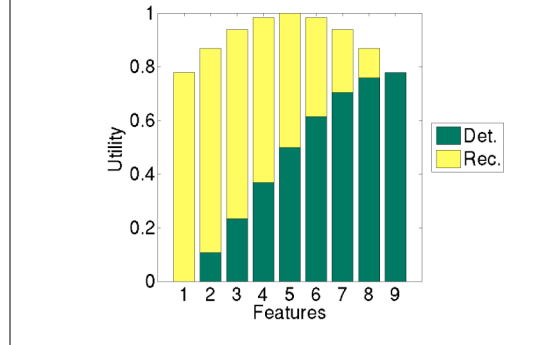
One Task, Many Problems



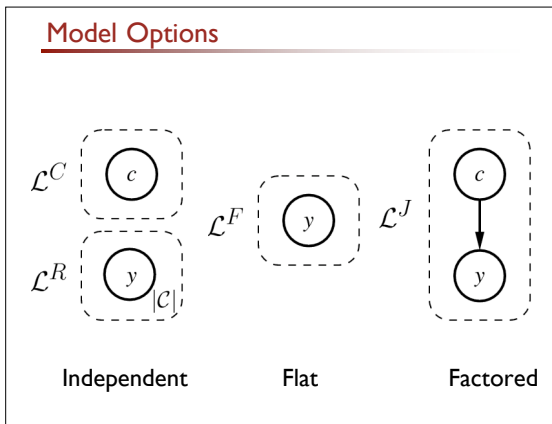
Reading Problem



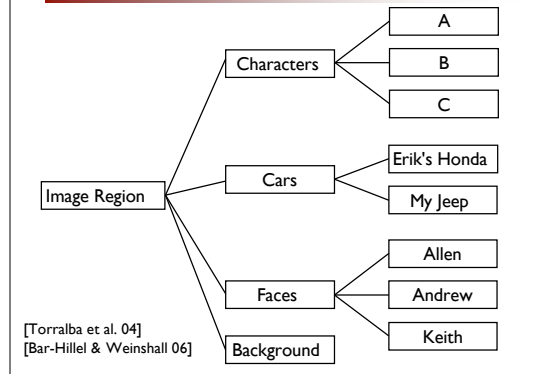
Which Features?



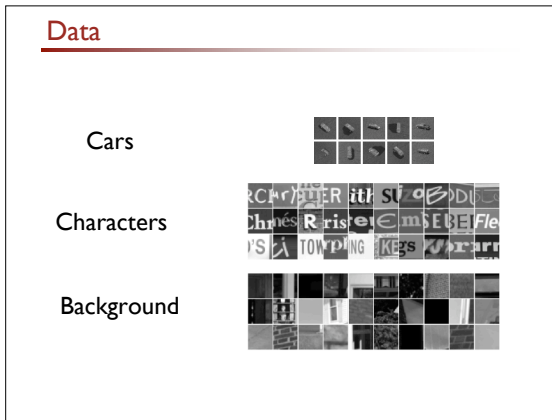
Model Options



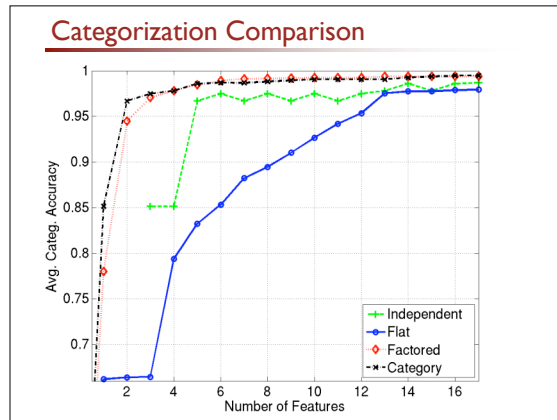
One Task, Many Problems



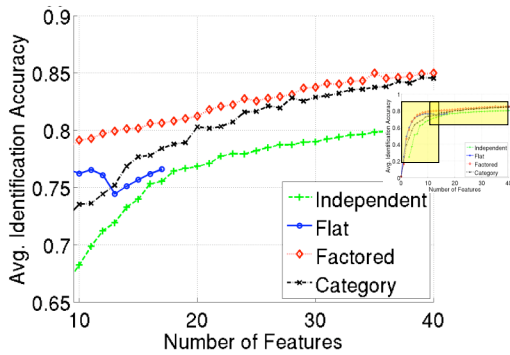
Data



Categorization Comparison



Recognition Comparison

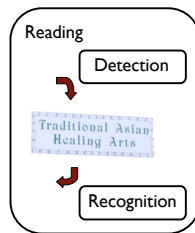


Contributions

- Better performance, less time
- Hierarchical training objective

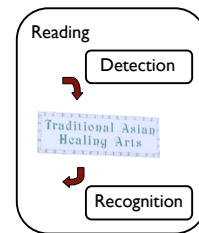
Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition



Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition

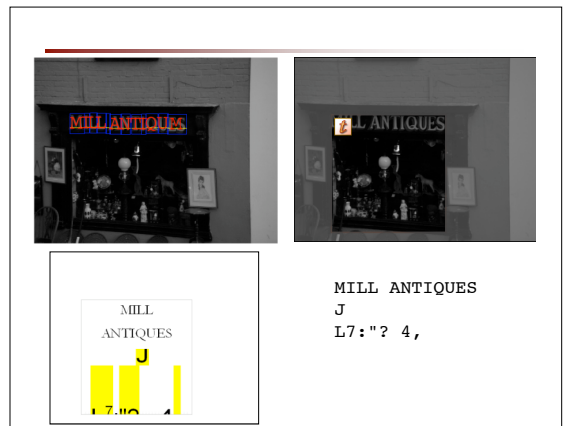
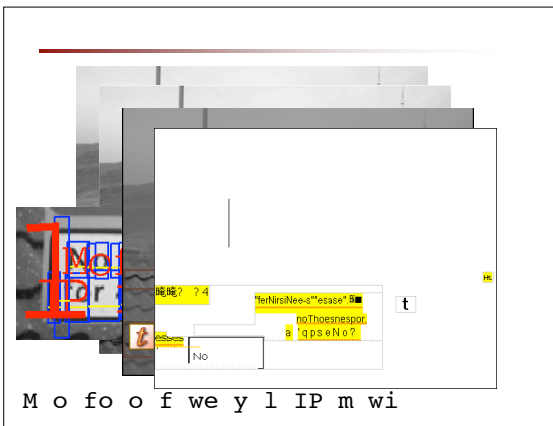
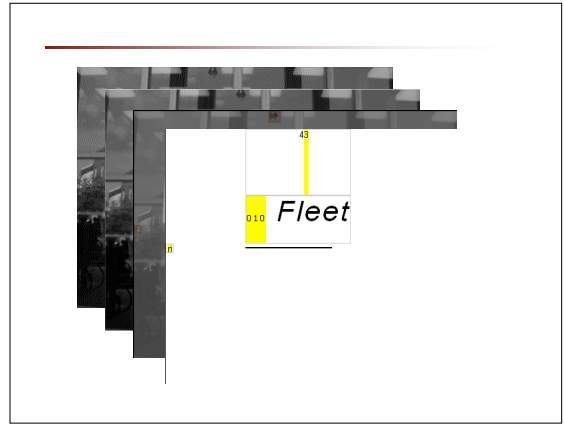


The Result?



OmniPage





Overview

- Detection
- Recognition
- Segmentation and Recognition
- Detection and Recognition

Reading

Detection

↪

Traditional Asian Healing Arts

↪

Recognition

Thank You

Support

NSF IIS-0326249
NSF IIS-0100851

NSA
CIA

Collaborators

Allen Hanson
Erik Learned-Miller
Andrew McCallum

Piyanuch Silapachote
Marwan Mattar
Richard Weiss