

FHT 2003

# Handwriting Recognition for Genealogical Records

Luke Hutchison  
lukeh@email.byu.edu

# Church Extraction Effort

- Nov 2002: Church released US 1880 and Canadian 1881 Census
- 55 million names
- **11 million man-hours**
- Granite Vault: contains 2.3 million rolls of microfilm  
( = about 6 million 300-page volumes )
- Approximate extraction time for one person  
(based on the above census): **280 years, 24/7**
- **We don't have that sort of time**
- Need automated extraction: handwriting recognition



# Example Microfilm Images

Edwin Lathay
Semina Lathay
William Lathay
Samuel Lathay
Mary Lathay
William Prigall
Samuel Fort
Walter Clark
Sarah Ann Clark
Walter Clark

Thomas Grafton	Newton
No. 45	
James Goodwin Quinn	The Hill Kumplian Park
No. 46	
John Millickap	The Pigeon House Upper Hill
No. 47	
Elizabeth Badham	The Villetts Upper Hill
No. 48	

Goff, Ira L.	Hans
— Gladys L.	Wife
— Emma L.	Daughter
Ruddy, Steve A.	Hans
— Melvina L.	Wife
— Henry J.	Son
— Rosa M.	Daughter
— Dora L.	Daughter
Hogallief, Anna J.	Hans
— Lora J.	Wife
— Mary E.	Daughter
— Vera J.	Daughter
— Anna L.	Son

Markir, Alma
— Elizabeth
— Wilcox
— Edward
Smith, Charles
— Alma
— Nellie C.
— Zia
Carlson, Emil
— Lillian
— Jean

Russel, Walter L.
—, Mary A.
—, Charles B.
Randolf, Clifton L.
—, Grace
—, Myrna
Sanders, Leibel R.
—, Bonnie C.
Mathis, Mattie L.
—, Harrietta B.
—, Henrietta D.
—, James L.

Shaw	Mary
Smade	Richard
Stone	
Smith	DECEASED
Spear	L.
Sedgwick	
Schiller	Samuel
Sullivan	Ellen
Still	Laura
Seiber	
Southy	
Sheedy	

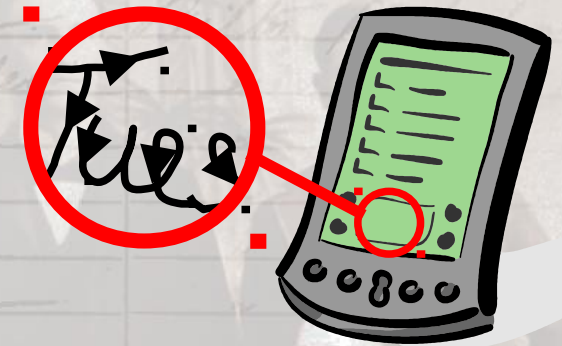


# Handwriting Recognition

- *Two different fields:*

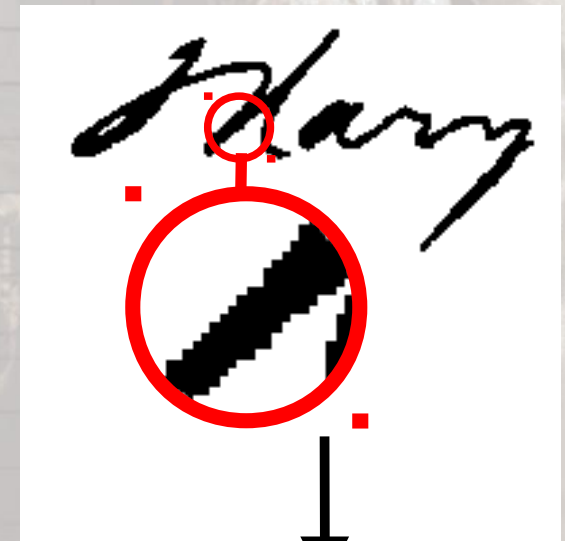
- **Online Handwriting Recognition**

- ☐ Writer's pen movements captured
- ☐ Velocity, acceleration, stroke order etc.
- ☐ Style can be constrained (e.g. Graffiti gestures)



- **Offline Handwriting Recognition**

- ☐ Only pixels
- ☐ Cannot constrain style (documents already written)

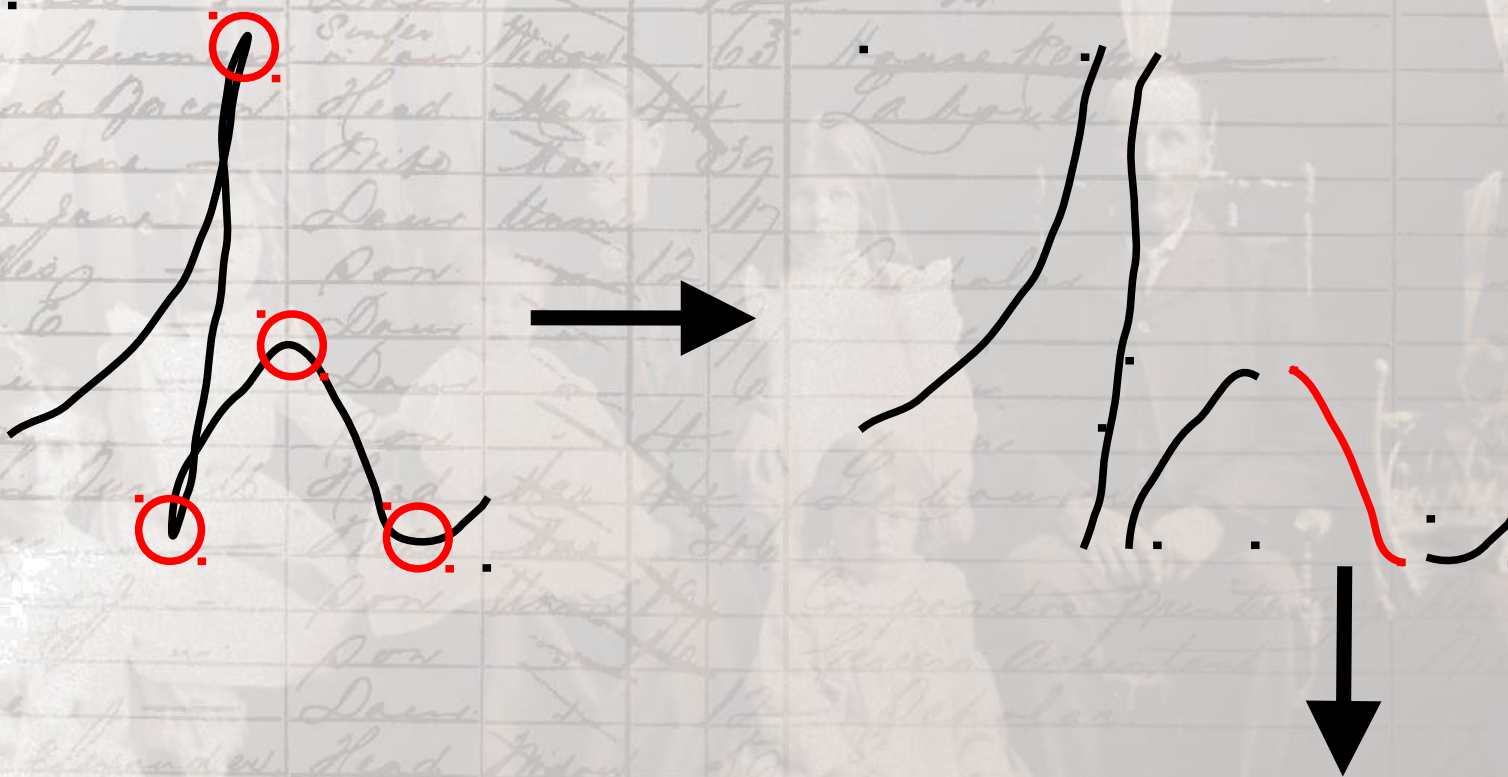


- *Offline is harder (less information)*
- *Genealogical records are all offline*

Mary

# Online Handwriting Recognition

- Modern systems are moderately successful,
- e.g. Microsoft Research's new Tablet PC:



Polynomial coefficients e.g. [0.94, 0.05, 0.29,...]

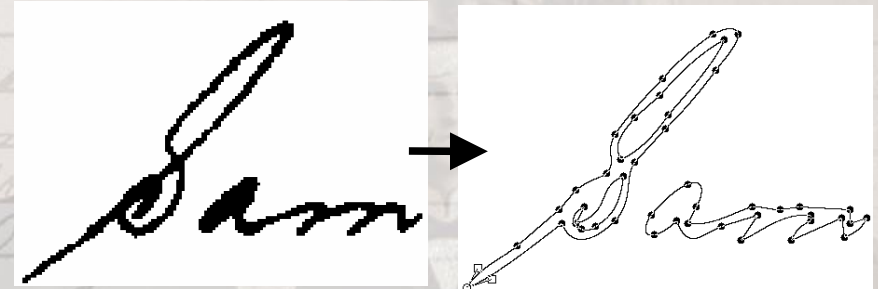


# Offline Handwriting Recognition

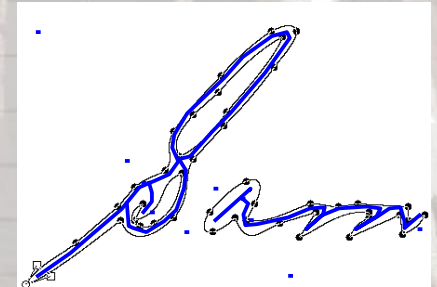
- *A difficult problem*
- *Almost as many approaches as there are researchers*
- *e.g.*
  - **Pattern Recognition**
  - **Statistical analysis**
  - **Mathematical modelling**
  - **Physics-based modelling**
  - **Subgraph matching / graph search**
  - **Neural networks / machine learning**
  - **Fractal image compression**
  - **... (too many to list) ...**

# Previous Work: Offline → Online Conversion

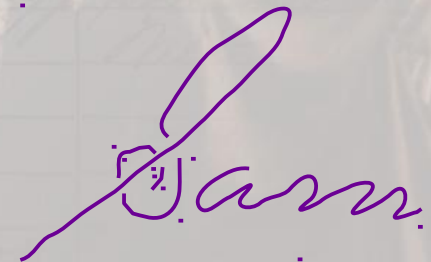
- Finding contour



- Finding midline



- Stroke ordering – difficult problem

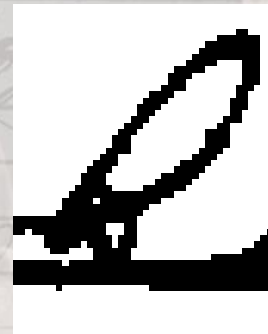




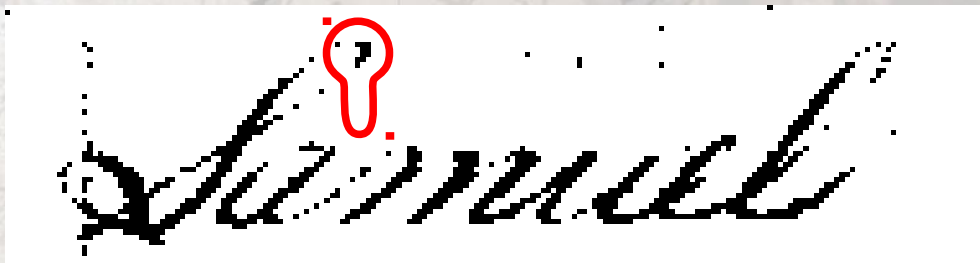
# Offline → Online Conversion ctd.

- Especially difficult with genealogical records:

- Stroke ordering: difficult



- Broken lines / blobs?



- Not practical



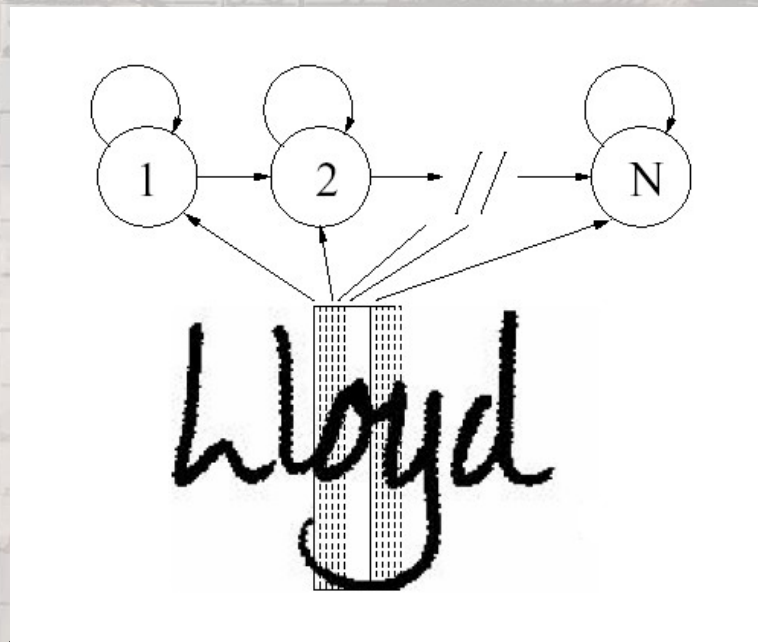
# Previous Work: Holistic Matching

- *Whole word is stretched to match known words*
- *Sources of variation compound across word*



# Previous Work: Sliding Window

- *Narrow vertical window slides across word*
- *A state machine recognizes sequences*

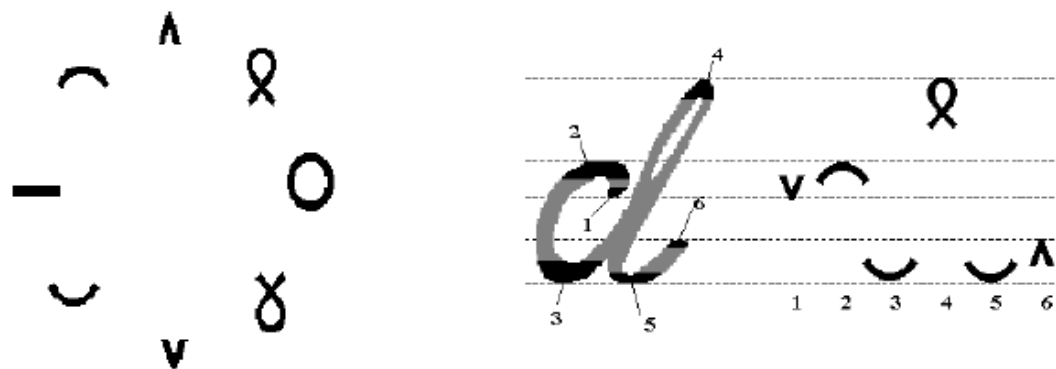


- *Results good, but sensitive to noise*



# Previous Work: Parascript

- Features detected & put in sequence
- Letters warped to best match sequence of features

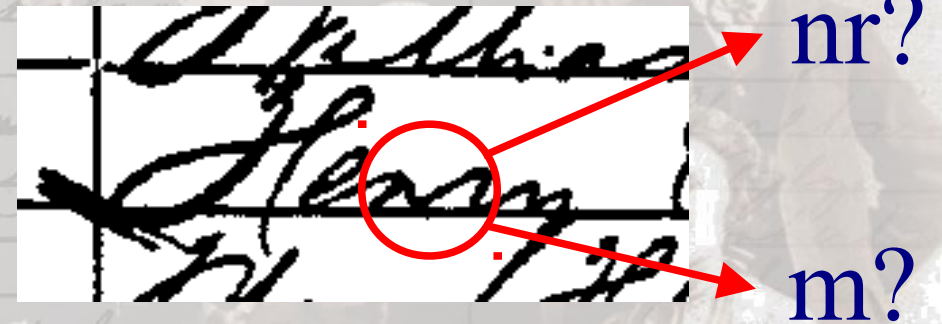


- Complex; sensitive to noise

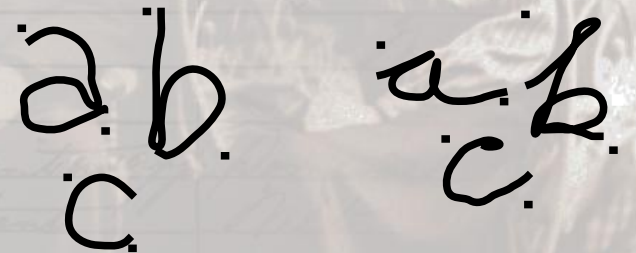
# Handwriting Recognition

- Some aspects of Handwriting Recognition:

- *Segmentation problem*  
(can't read word until it is segmented; can't segment word until it is read)



- *Different handwriting styles*



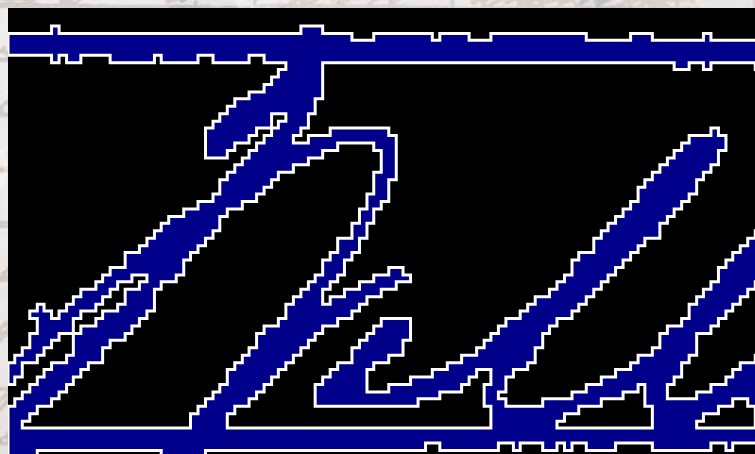
- *Use of dictionary to correct for errors in reading*

Srnitb --> Smith



# Thesis Approach: Preprocessing

*Outlines of word are traced and smoothed:*

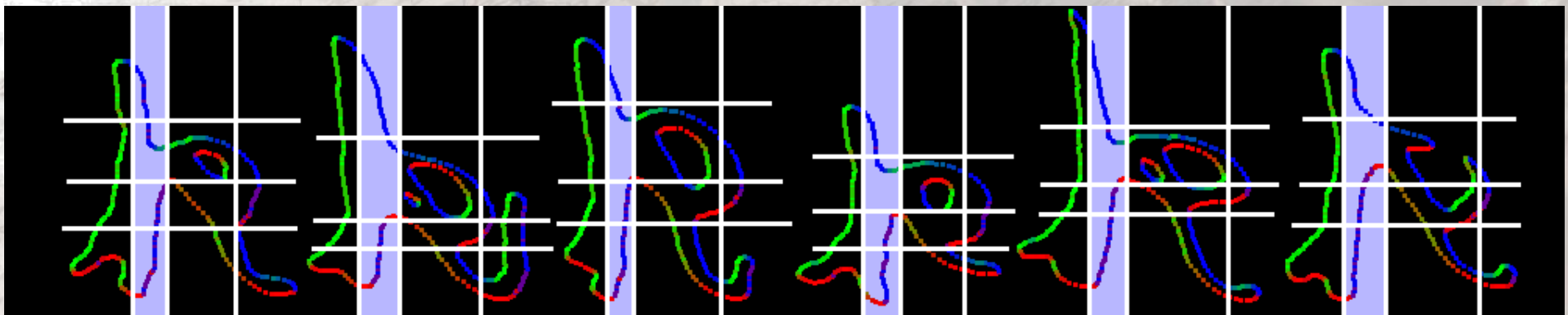


*Handwriting slope is corrected for automatically:*



# Segmentation

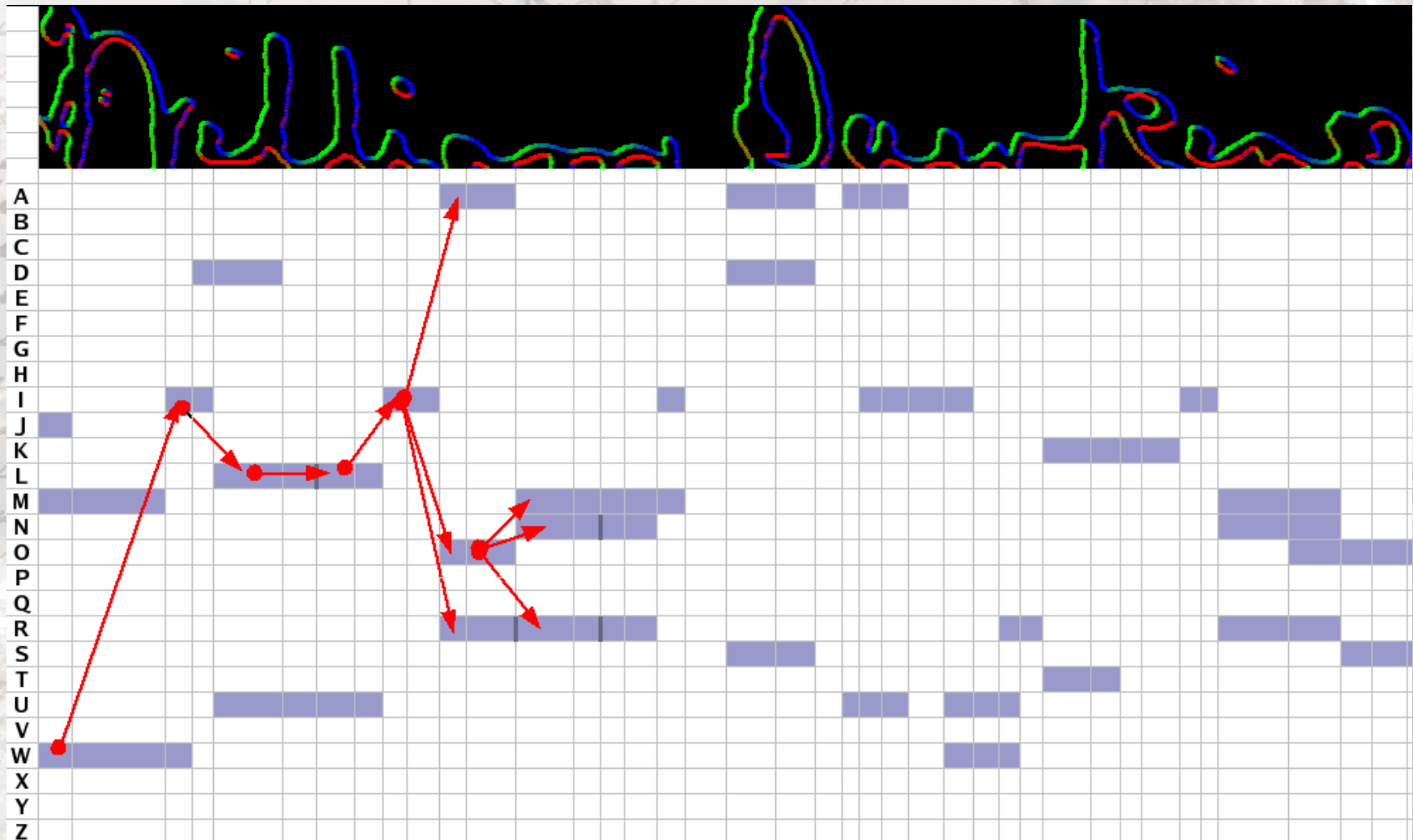
- Goal: robustly cut letters into segments
- Match multiple segments to detect letters
- Easier than matching whole letter





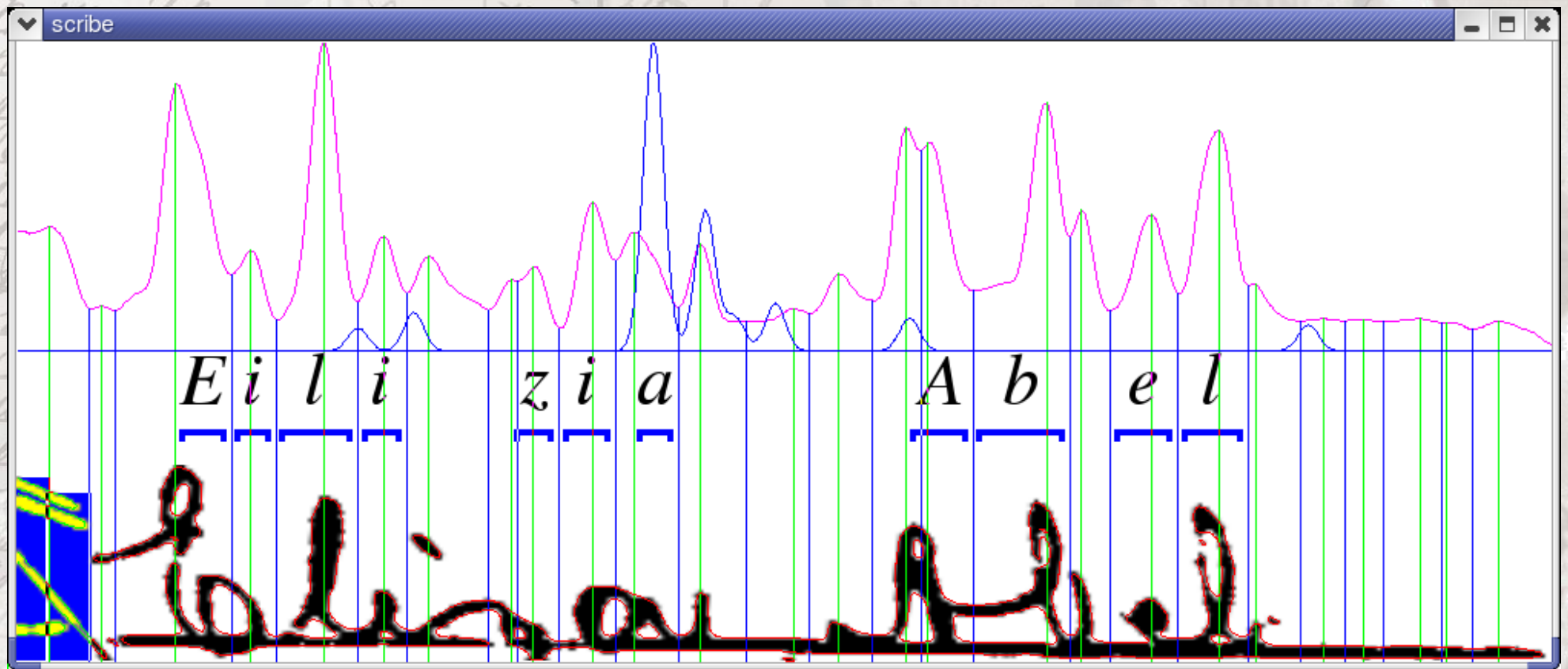
# Dynamic Global Search

- Assemble word spelling from possible letter readings



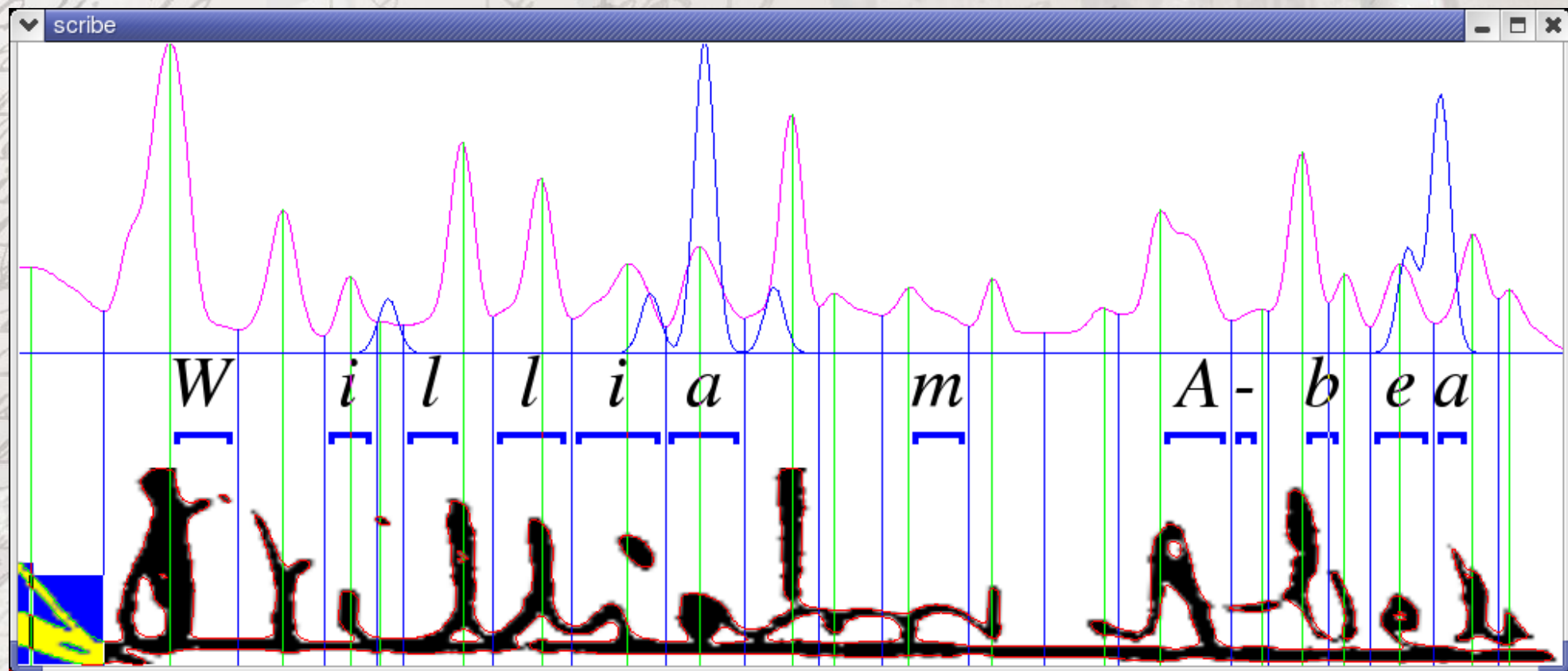
Best path: "Williarw Suwkino" (65% confidence)

# Results (1)

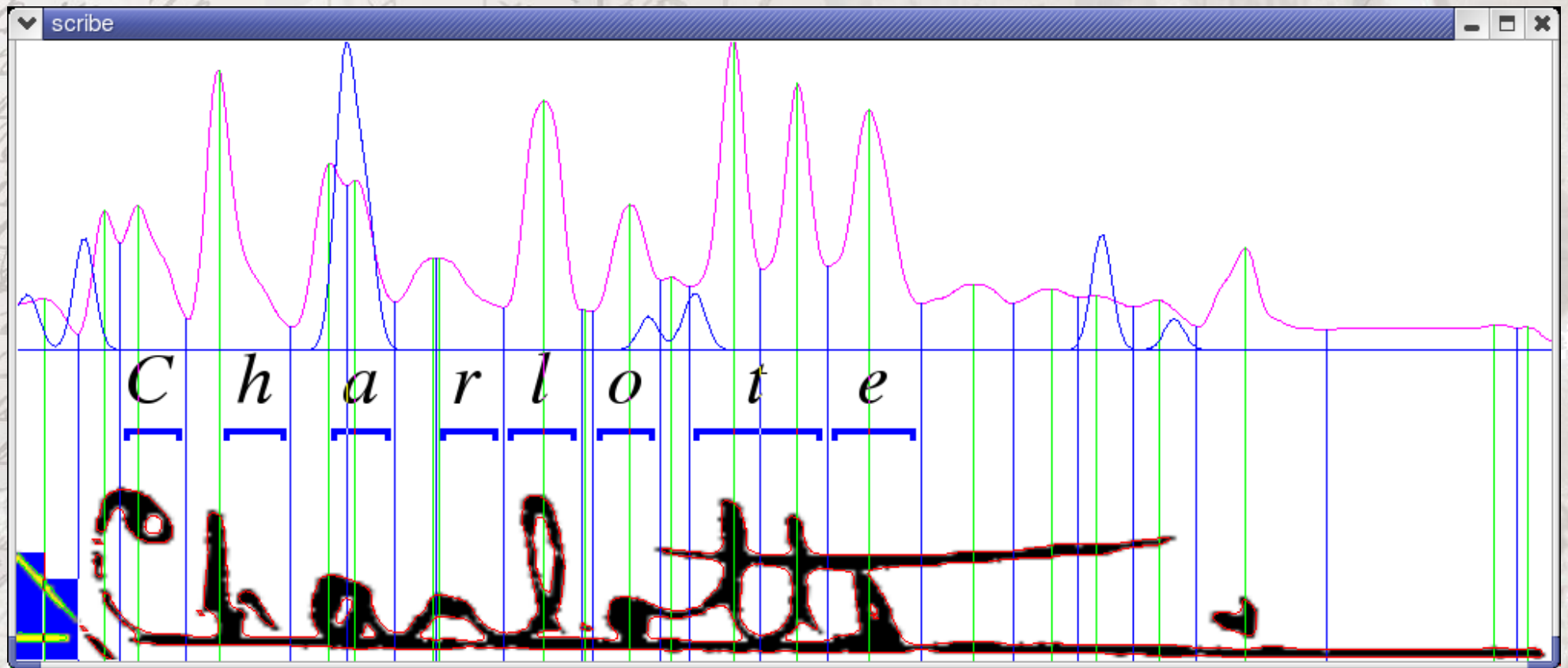




# Results (2)

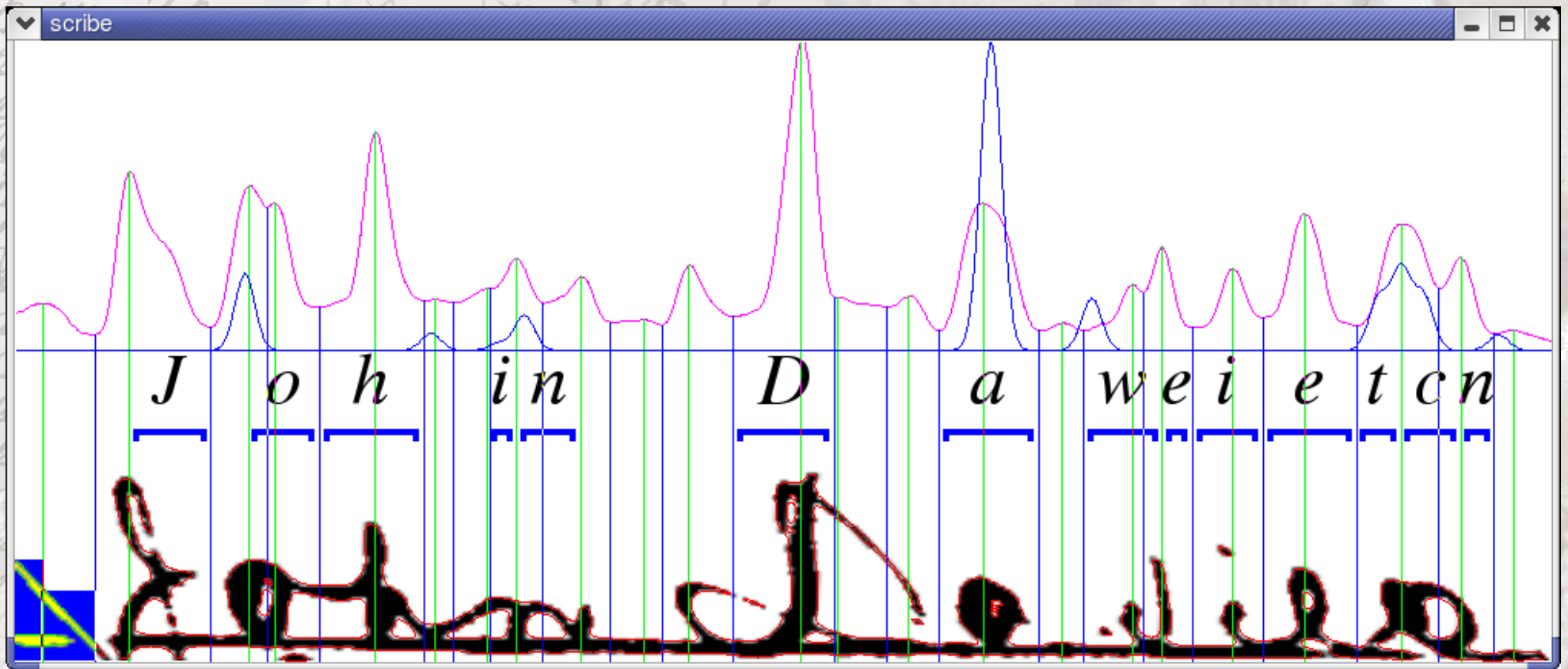


# Results (3)



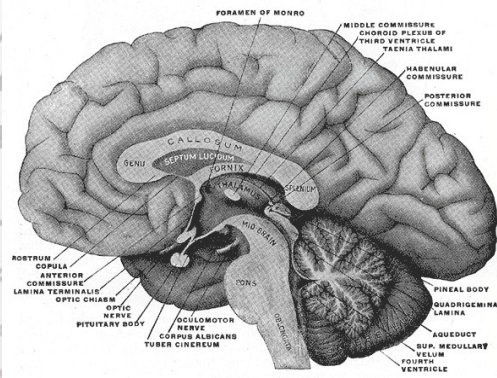


# Results (4)



In general: results even worse – system only worked well on words it was specifically trained on

# The Human Brain's Visual System

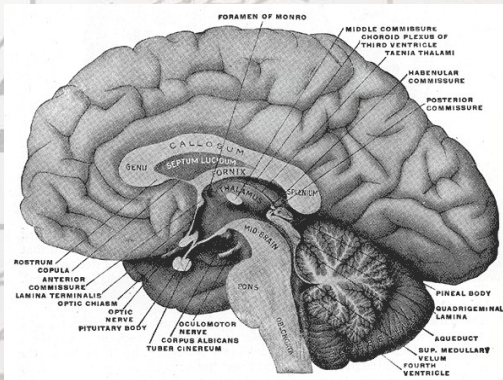


Retina



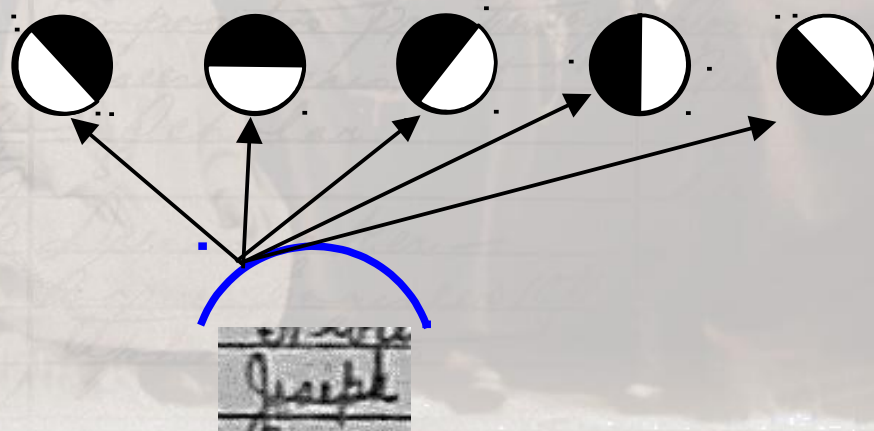


# The Human Brain's Visual System

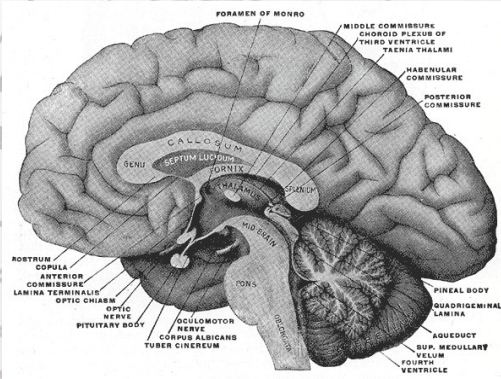


Angular edge detectors

Retina



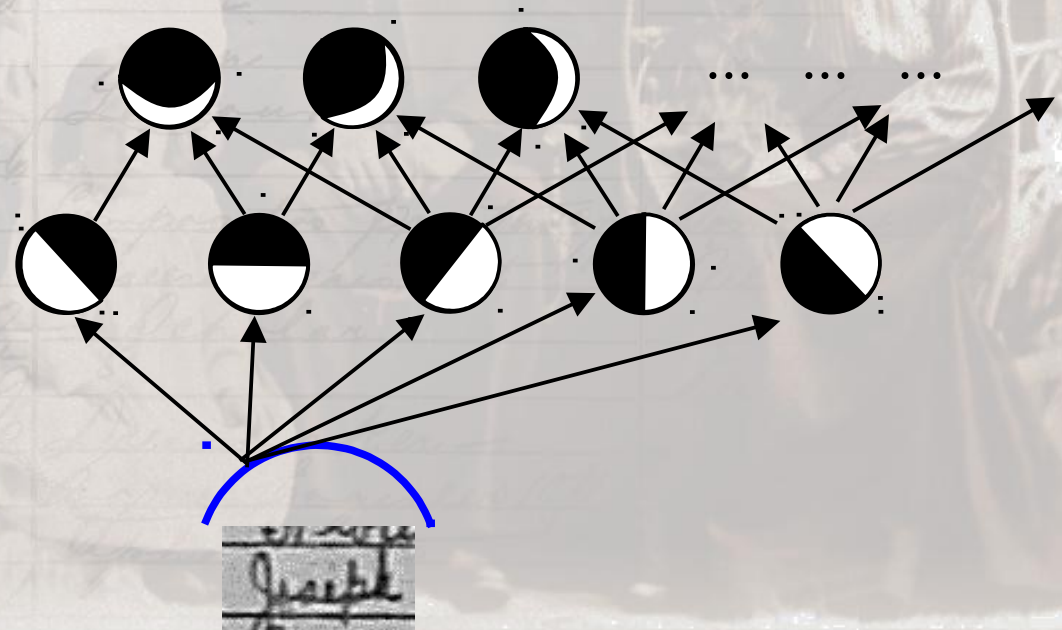
# The Human Brain's Visual System



Line / curve detectors

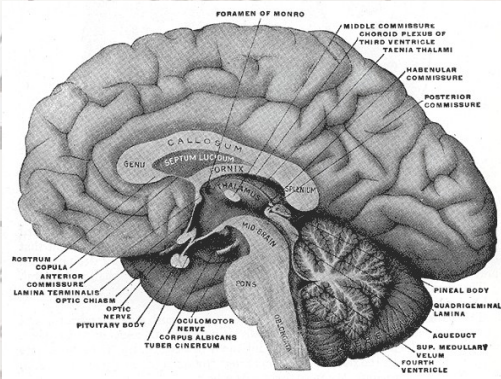
Angular edge detectors

Retina





# The Human Brain's Visual System

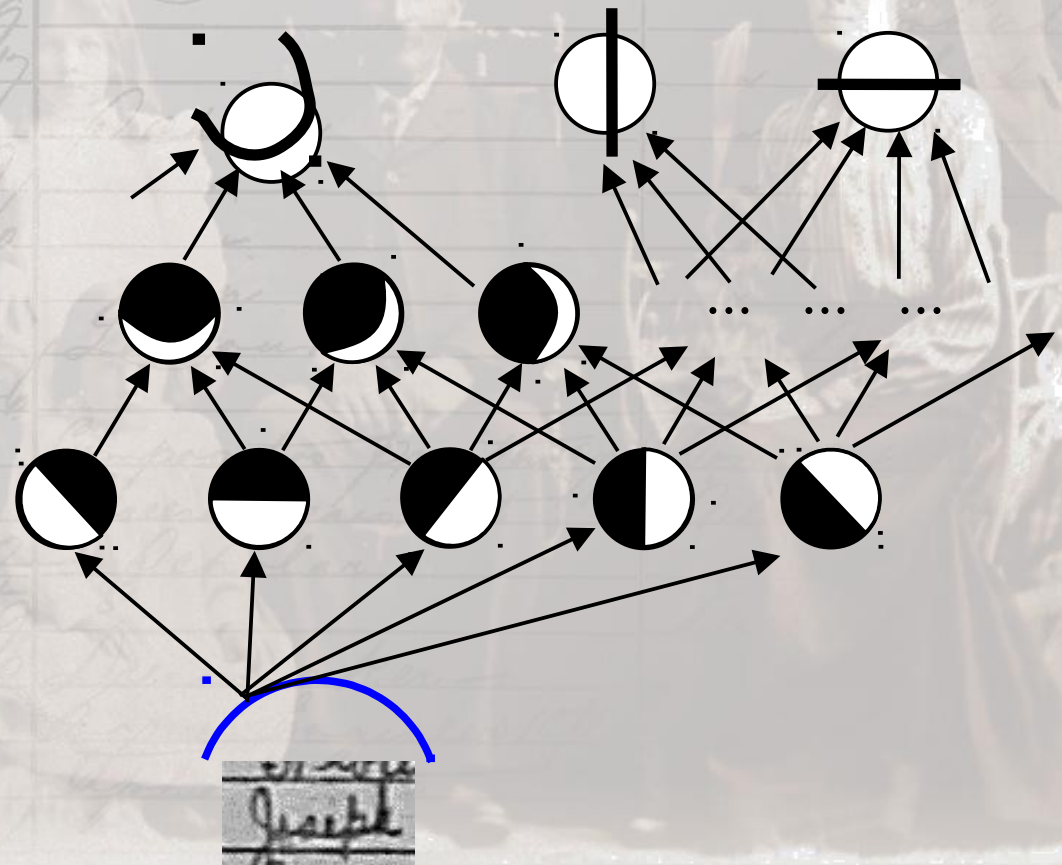


Feature detectors

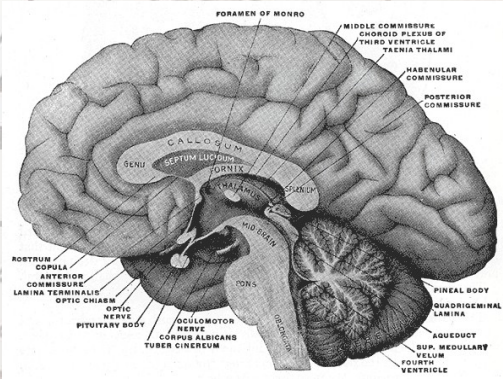
Line / curve detectors

Angular edge detectors

Retina



# The Human Brain's Visual System

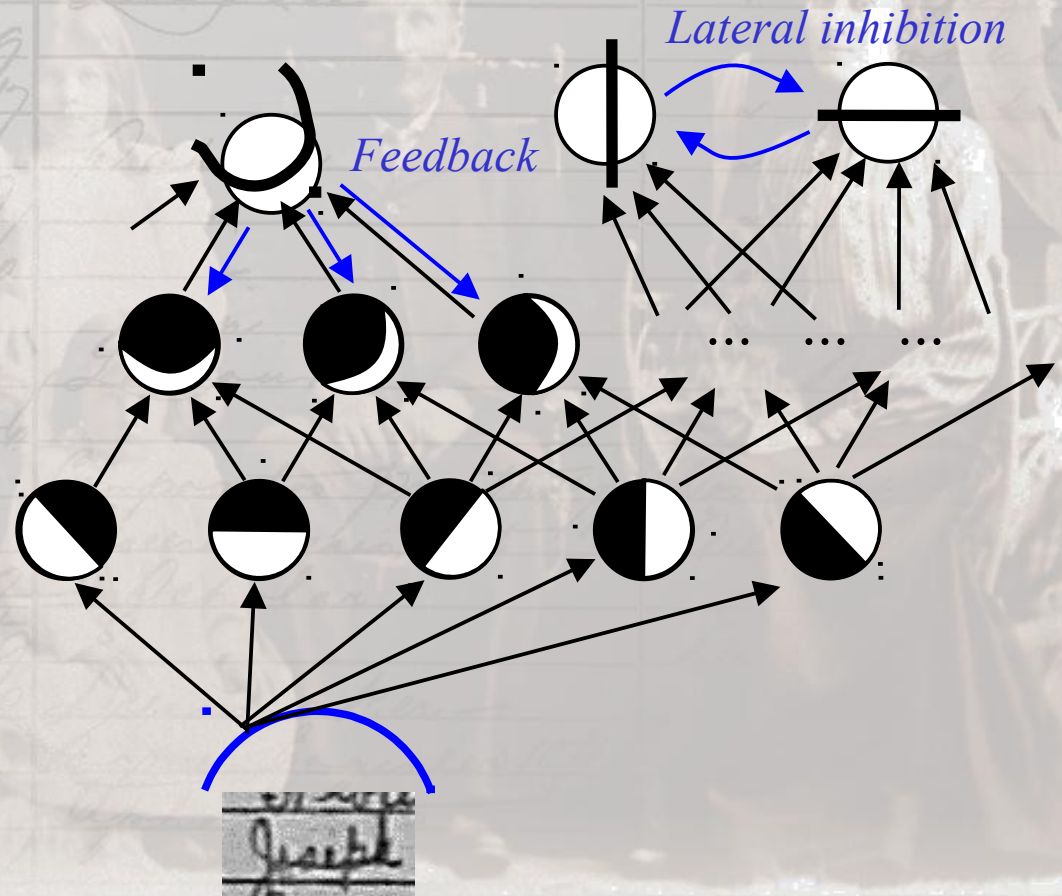


Feature detectors

Line / curve detectors

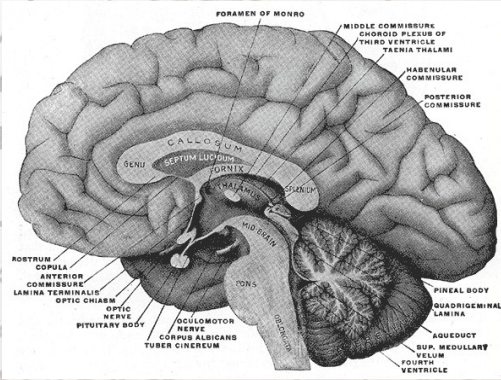
Angular edge detectors

Retina





# The Human Brain's Visual System



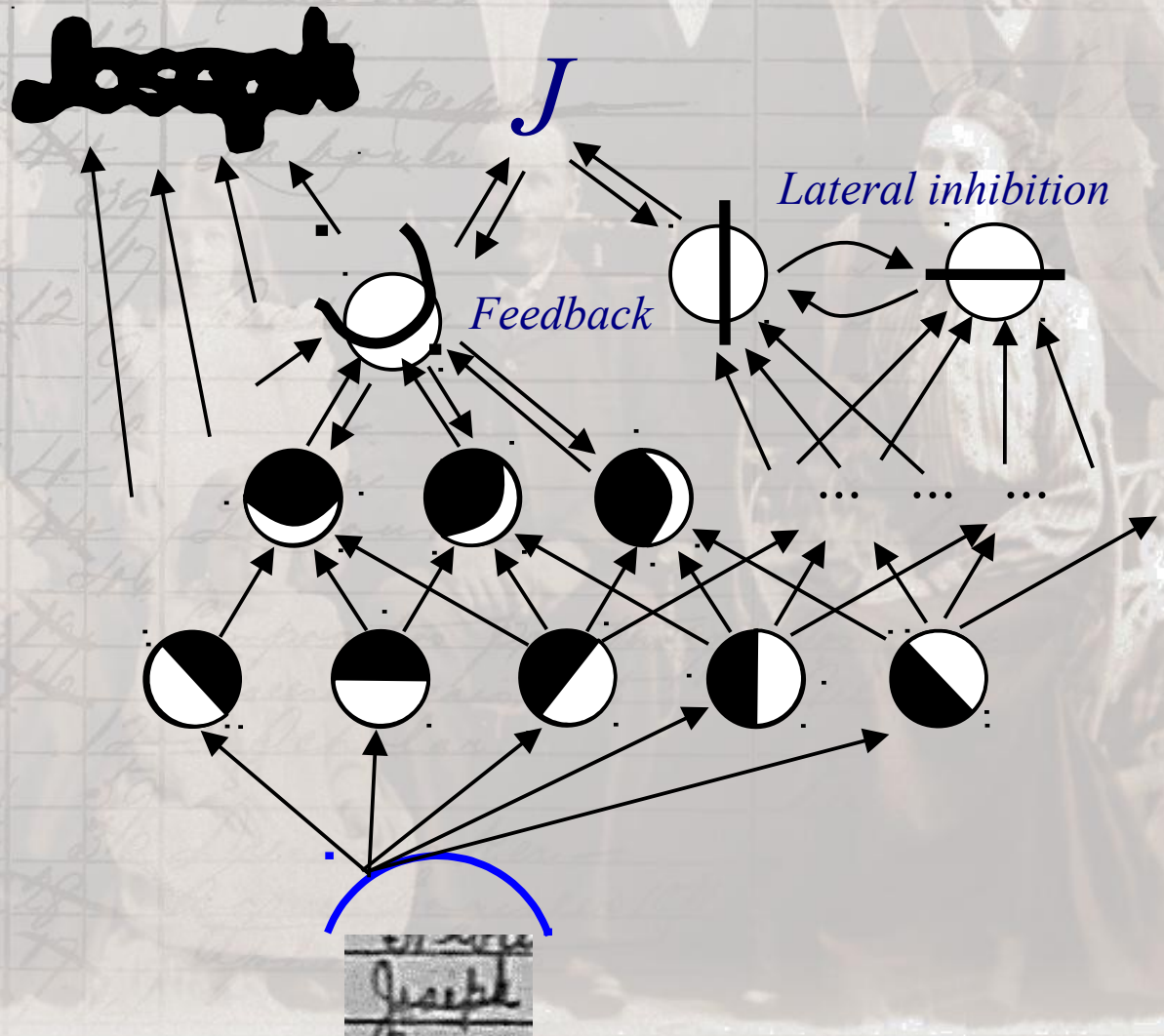
Letter / word shape  
recognizers

Feature detectors

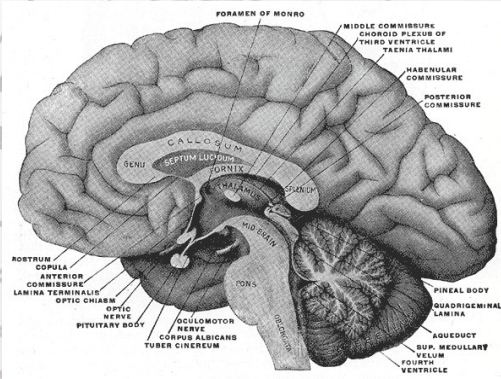
Line / curve detectors

Angular edge detectors

Retina



# The Human Brain's Visual System



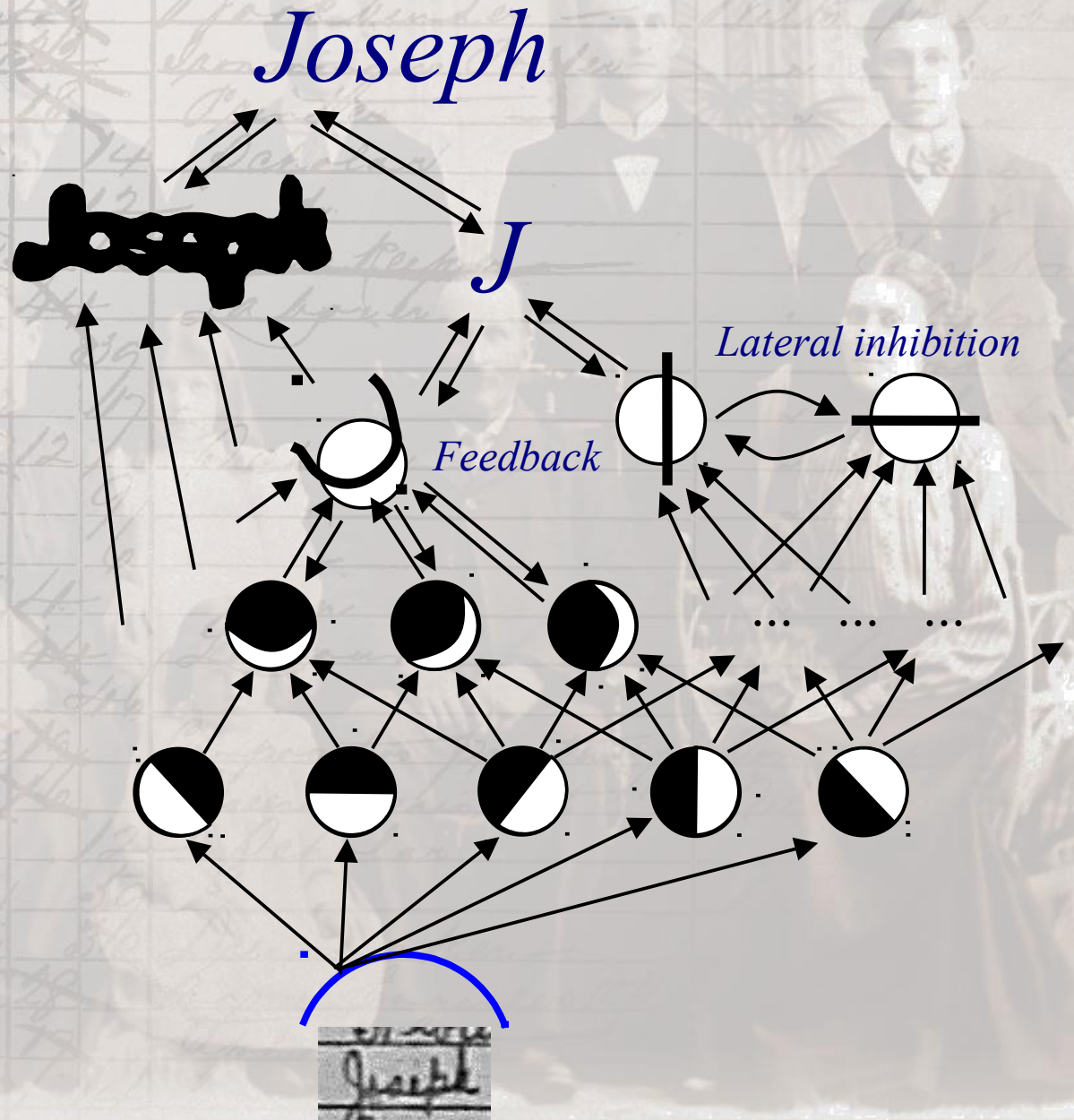
Letter / word shape  
recognizers

Feature detectors

Line / curve detectors

Angular edge detectors

Retina





# Conclusions

- *Handwriting recognition is important for genealogy...  
...but it is hard*
- *Current methods don't work very well...  
...and they don't operate much like the human brain*
- *Future work should focus on understanding the brain, and emulating it as much as possible, e.g. With:*
  - **Hierarchical reasoning**
  - **Feedback**
  - **Lateral inhibition**

Rough of		Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of		Urban Sanitary District of	Rural Sanitary District of	Electoral
Salisbury		St Edmund	Salisbury			Salisbury		
No. of Inhabited, or (if less than 10) of the	NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of		Rank, Profession, or OCCUPATION	WHERE BORN	
				Male	Female			
	William Spender	Head	Married	49		Bookbinder	Wiltshire Salisbury	
	William J. a.	Son	Unmarried	18		Spencerian Penman	Salisbury	
	Edward a.	Son	Unmarried	16		Penman	Salisbury	
	Gerard a.	Son	Unmarried	14		Penman	Salisbury	
	Alfred a.	Son	Unmarried	12		Penman	Salisbury	
	Maria Spender	Wife	Married	63		Housekeeper	Salisbury	
	Richard Jacob	Head	Married	44		Labourer	Salisbury	
	Emma Jane	Wife	Married	39			Salisbury	
	Emma Jane	Son	Unmarried	17			Salisbury	
	Charles	Son	Unmarried	12			Salisbury	
	Alfred	Son	Unmarried	9			Salisbury	
	Alfred	Son	Unmarried	6			Salisbury	
	Harry	Son	Unmarried	4			Salisbury	
	Edmund Spender	Head	Married	44		Labourer	Salisbury	
	Edmund	Wife	Married	44			Salisbury	
	Edmund	Son	Unmarried	16		Compositor Printer	Salisbury	
	Edmund	Son	Unmarried	14		Printer	Salisbury	
	Edmund	Son	Unmarried	12			Salisbury	
	Edmund	Head	Married	50			Salisbury	
	Edmund	Son	Unmarried	18			Salisbury	
	Edmund	Son	Unmarried	16			Salisbury	
	Edmund	Son	Unmarried	14			Salisbury	
	Edmund	Son	Unmarried	12			Salisbury	
	Edmund	Son	Unmarried	10			Salisbury	
	Edmund	Son	Unmarried	8			Salisbury	
	Edmund	Son	Unmarried	6			Salisbury	
	Edmund	Son	Unmarried	4			Salisbury	

Questions?

Luke Hutchison  
lukeh@email.byu.edu



No. of inhabited houses are situated within the boundaries of the							
Rough of	Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of		Urban Sanitary District of	Rural Sanitary District of	
Salisbury	St Edmund	Salisbury	Salisbury		Salisbury		
No. of Inhabited Houses or Flats	NAME and Surname of each Person	RELATION to Head of Family	CON- DITION as to Marriage	AGE last Birthday of		Rank, Profession, or OCCUPATION	WHERE BORN
				Male	Female		
	William Appender	Head	Married	49		Bookbinder	Wiltshire Salisbury
	William D. A.	Son	Unmarried	18		Spencerian Penman	Salisbury
	Edward A.	Son	Unmarried	18		Printer	Salisbury
	Gerard A.	Son	Unmarried	14		Scholar	Salisbury
	Alfred A.	Son	Unmarried	12		do.	Salisbury
	Maria Newman	Sister in Law	Married	63		Housekeeper	Salisbury
	Richard Jacob	Head	Married	44		Labourer	Salisbury
	Emma Jane	Wife	Married	39			Salisbury
	Emma Jane	Son	Unmarried	17			Salisbury
	Charles	Son	Unmarried	12		Scholar	Salisbury
	Alfred	Son	Unmarried	9			Salisbury
	Alfred	Son	Unmarried	6			Salisbury
	Harry	Son	Unmarried	4			Salisbury
	Edmund Bennett	Head	Married	44		Labourer	Salisbury
	Edmund	Wife	Married	44			Salisbury
	Edmund	Son	Unmarried	14		Compositor Printer	Salisbury
	Edmund	Son	Unmarried	14		Grammar Assistant	Salisbury
	Edmund	Son	Unmarried	12		Scholar	Salisbury
	Edmund Appender	Head	Married	55			Salisbury
	Edmund Appender	Son	Unmarried	36		Printing Pressman	Salisbury
	Edmund Appender	Son	Unmarried	28		Printer	Salisbury
	Edmund Appender	Son	Unmarried	17		Printer	Salisbury
	Edmund Appender	Son	Unmarried	14			Salisbury