

*LUKE A. D. HUTCHISON  
(lukeh@byu.net)*

# Document Image Registration

## Example: Data Extraction (UDE)

UDE - [Packet - C:\UDEWIN\DATA\A10101-1.GGD - 1 G Sarah/Britton F ]

File Edit Batch Packet Records Fields Image Reports Search Window Help

**lloyd** d. Boat No. 1  
(If rural, give location)

Last) 4. Date (Month) (Day) (Year)  
ith 4. Date (Month) (Day) (Year)  
Death May 28, 1953

9. Age (In years If Under 1 Year If Under 24 Hrs.  
last birthday) Months Days Hours Min.  
13 7 45

11. State and county or 12. Citizen of What Country?  
Alabama U.S.A.

PRINCIPAL'S NAME: Sarah/Britton  
Sex: F  
Death Date: 28May1953  
Death Place: Birmingham, Jefferson County, Alabama  
Death Date: 28May1953 Data Entry Mode

Press Shift+F1 for keyboard shortcuts or Shift+F2 for diacritics keys

UDE - [Packet - C:\UDEWIN\DATA\A10101-1.GGD - 1 G Abraham/Fairbanks M ]

File Edit Batch Packet Records Fields Image Reports Search Window Help

(If rural, give location)  
12  
Last) 4. Date (Month) (Day) (Year)  
Death 4. Date (Month) (Day) (Year)  
9 - 26 - 53

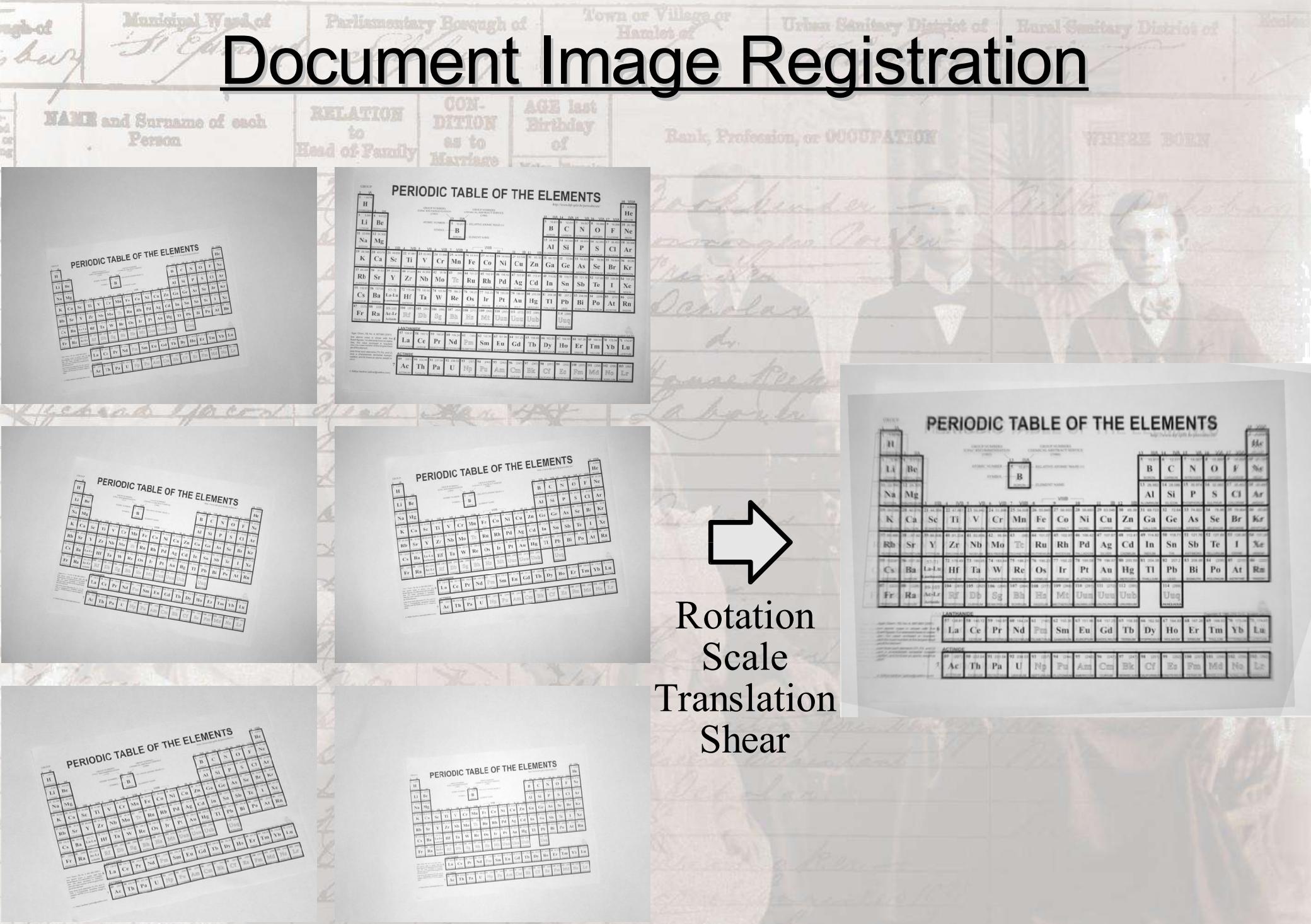
9. Age (In years If Under 1 Year If Under 24 Hrs.  
last birthday) Months Days Hours Min.  
12 22 0 0 0 0

11. State and county or 12. Citizen of What Country?  
Ala U.S.A.

PRINCIPAL'S NAME: Abraham/Fairbanks  
Sex: M  
Death Date: 26Sep1953  
Death Place: Birmingham, Jefferson County, Alabama  
Death Date: 26Sep1953 Data Entry Mode

Restore field to previous state

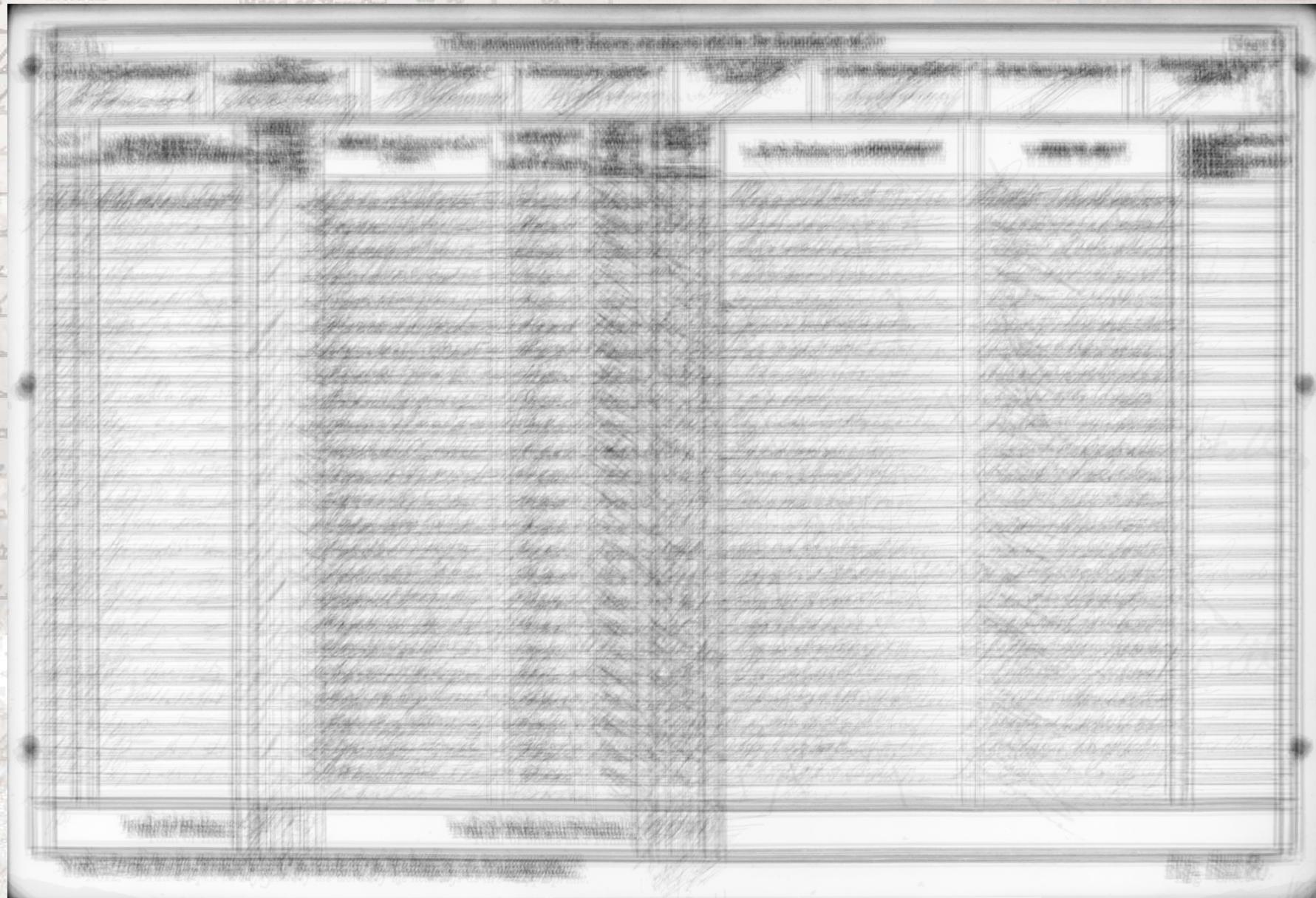
# Document Image Registration



The boundaries of the Local Government Districts are defined within the boundaries of the					
Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of	Urban Sanitary District of	Rural Sanitary District of	Electoral
NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of	Bank, Profession, or OCCUPATION	WHERE BORN
			Years Females		
William Pendleton	Head	Married		Bookbinder	Bethel Pendleton
William D.	Son	Married		Engineering Officer	
Edward	Son	Married		Clerk	
Georgie	Son	14		Scholar	
Salmon	Son	13			
Maria Freeman	Sister	Never Married		Housekeeper	
Richard Freeman	Head	Never Married			
James Jane	Daughter	10			
James Jane	Daughter	17			
Charles	Son	12		Scholar	
Alberto	Son	9			
Annie	Son	6			
Harry	Son	4			
John Pendleton	Head	Married		Labourer	
Eliza	Daughter	8			
John	Son	10		Compositor	
John	Son	12		General Labourer	
Albert Pendleton	Son	12		Scholar	
Matthew Pendleton	Head	Married	50		
John Pendleton	Son	32		Labourer	
John Pendleton	Son	18			
John Pendleton	Son	77			

# [Animations]

# Document Image Registration



Before registration (pointwise mean of 27 unregistered images)

# Document Image Registration

**NOTE.**—Draw the pen through such of the words of the headings as are inappropriate

Eng- Sheet G

After registration (pointwise mean of registered images)

# Document Image Registration

**NOTE.**—Draw the pen through such of the words of the headings as are inappropriate.

Pointwise median of registered images, producing a blank form

# Document Image Registration

Municipal Ward of		Parliamentary Borough of		Town or Village or Hamlet of		Urban Sanitary District of		Rural Sanitary District of	
<i>St Edmund</i>		<i>Salisbury</i>		<i>Salisbury</i>		<i>Salisbury</i>		<i>Salisbury</i>	
NAME and Surname of each Person		RELATION to Head of Family		CON-DITION as to Marriage		AGE last Birthday of Male Females		Bank, Profession, or OCCUPATION	
<i>William Pender Head</i>		<i>William Pender Head</i>		<i>Single</i>		<i>21</i>		<i>Bookbinder</i>	
<i>William Pender Son</i>		<i>William Pender Son</i>		<i>Single</i>		<i>20</i>		<i>Bookbinder</i>	
<i>City or Municipal Borough of</i>		<i>Municipal Ward of</i>		<i>Parliamentary</i>				<i>WHERE BORN</i>	
<i>Salisbury</i>		<i>St Edmund</i>		<i>Salisbury</i>				<i>Salisbury</i>	
<b>HOUSES</b>									
In-habit-ed	Un-habited (U.) or Building (B.)	NAME and Surname of each Person	RELATION to Head of Family						
1		<i>James Parsons</i>	<i>Head</i>						
		<i>Annie</i>	<i>Daughter</i>						
		<i>Alfred</i>	<i>Son</i>						
		<i>Edith</i>	<i>Daughter</i>						
		<i>Ernest</i>	<i>Son</i>						
		<i>Elizabeth Harrington</i>	<i>Daughter</i>						

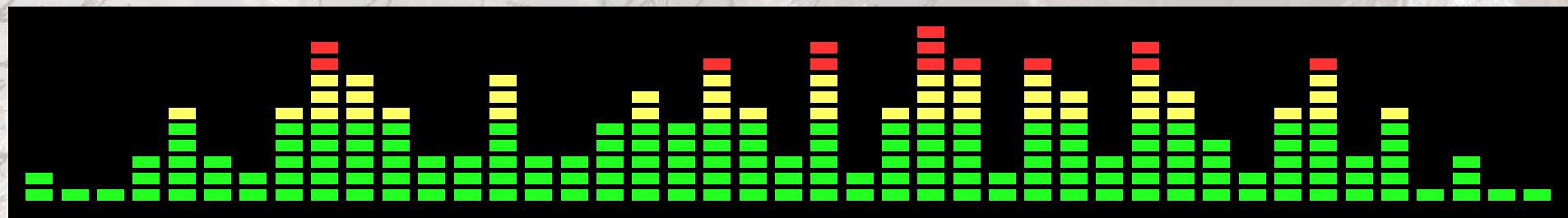
Predictive image compression by registration (14% reduction)

# The Discrete Fourier Transform (DFT)

- Conversion from time domain to frequency domain



← Time →

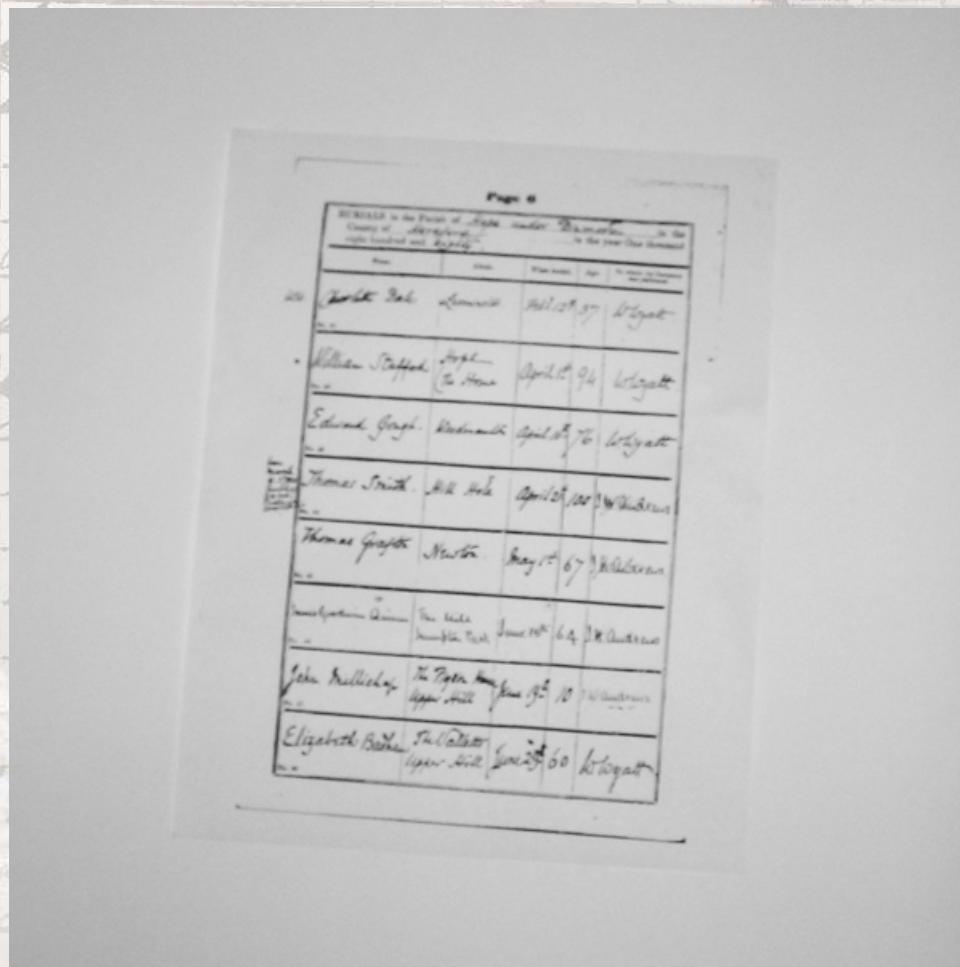


← Frequency →

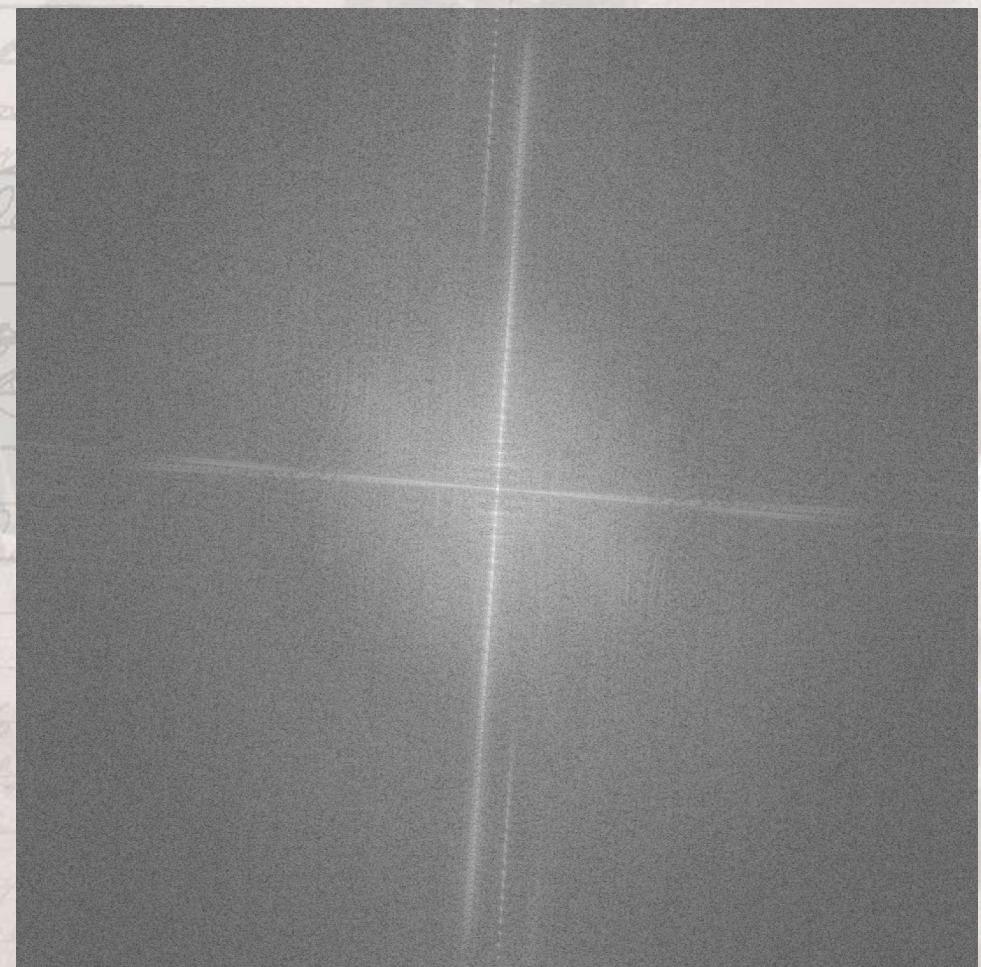
- Can also perform DFT in 2 or more dimensions:  
breaks down spatially-distributed signal (e.g. image)  
into individual frequency components

borough of Salisbury	Municipal Ward of <i>St Edmund</i>	Parliamentary Borough of <i>Salisbury</i>	Town or Village or Hamlet of	Urban Sanitary District of <i>Salisbury</i>	Rural Sanitary District of <i>Salisbury</i>
NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	of Males Females	Rank, Profession, or OCCUPATION	WHERE BORN

# The 2D DFT

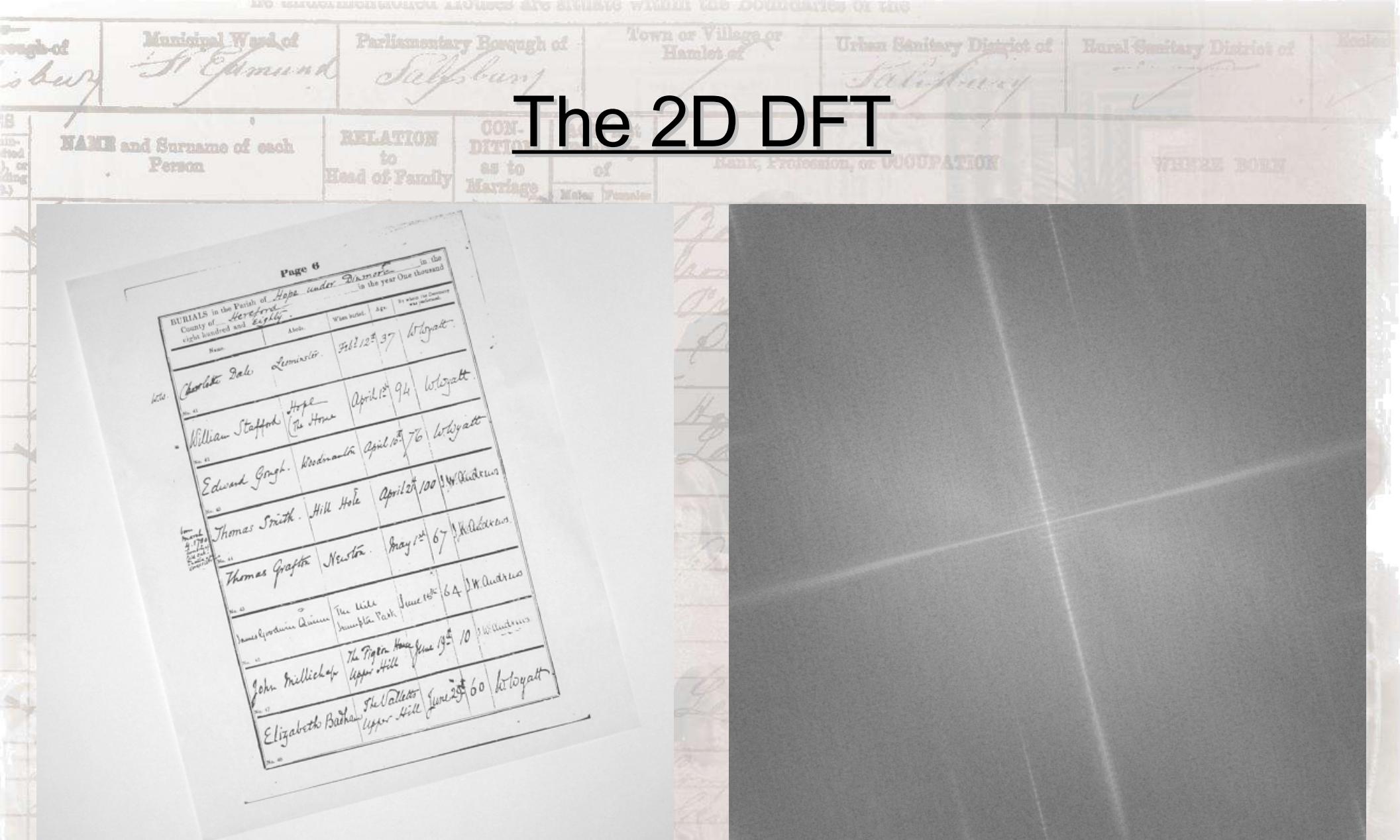


Document image



DFT of document image

- Sets of parallel lines act as periodic “wavefronts”



Document image

DFT of document image

- Sets of parallel lines act as periodic “wavefronts”

# The Mellin Transform

- A function is sampled at exponential intervals to produce a log-domain representation

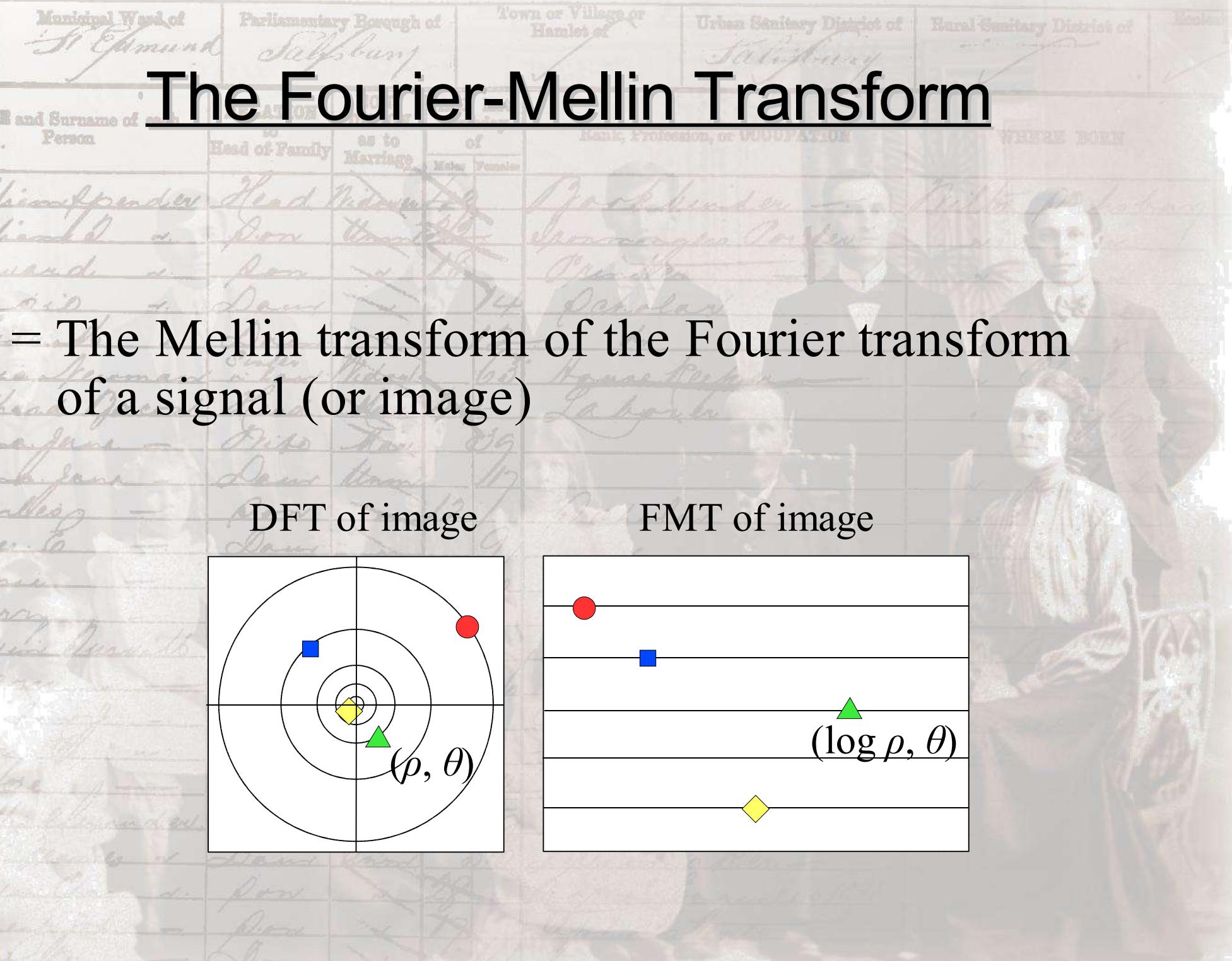
$$\phi(z) = \int_0^\infty t^{z-1} f(t) dt$$

- Scaling a function by a factor of  $S$  results in a shift in the Mellin transform by  $\log(S)$ :

$$x' = x \cdot S$$

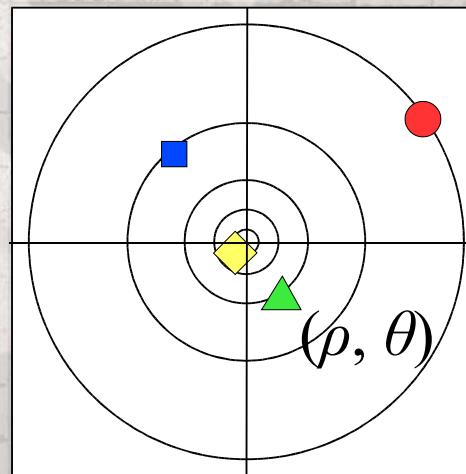
$$\log(x') = \log(x) + \log(S)$$

# The Fourier-Mellin Transform

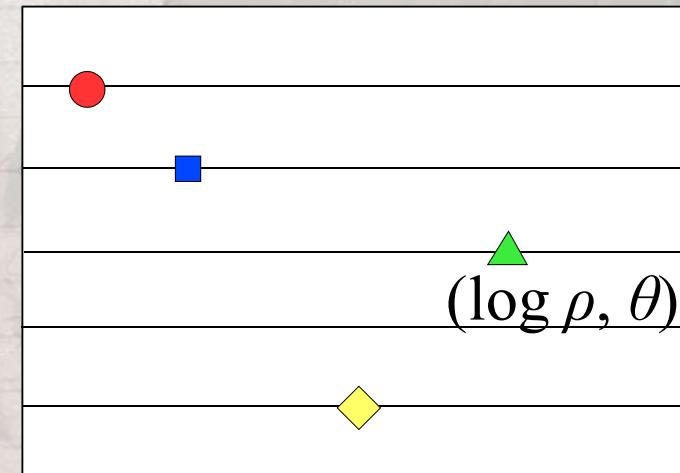


= The Mellin transform of the Fourier transform of a signal (or image)

DFT of image

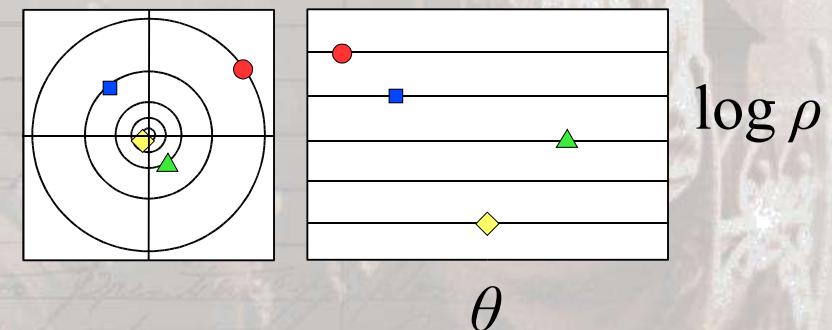


FMT of image



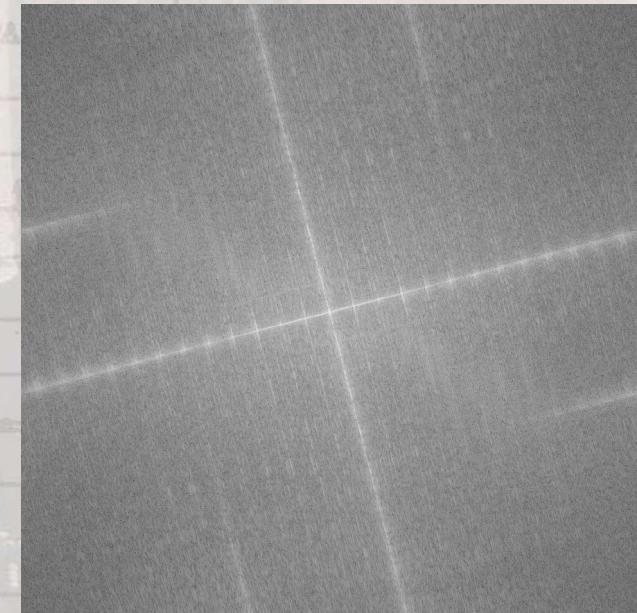
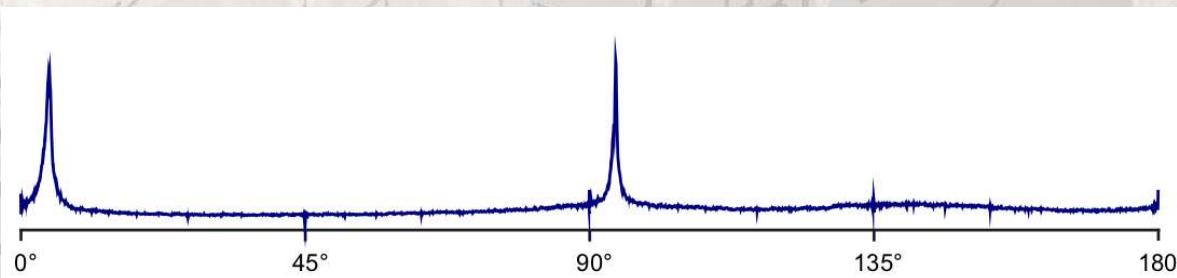
# Fourier-Mellin Transform for Image Registration

- Fourier Shift Theorem => ignore translation when dealing with magnitudes
- Mellin Transform of DFT used for registration
  - $(\log \rho, \theta)$  histogram: Recovery of scale, rotation is a 2D correlation problem
- Several issues, e.g.
  - Rotationally-dependent aliasing
  - “Edge Effects” (rotation and periodicity don’t commute)

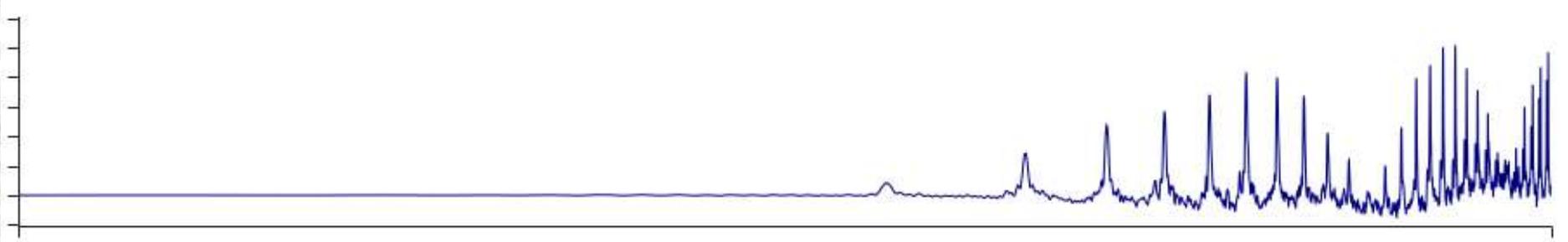


# Parameter Recovery

- Rotation [+ shear]: Recover document axes by finding two largest peaks in radial projection histogram



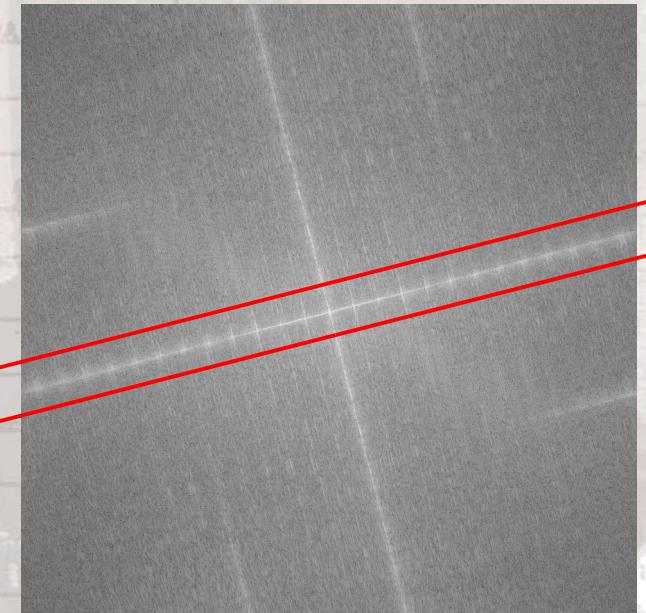
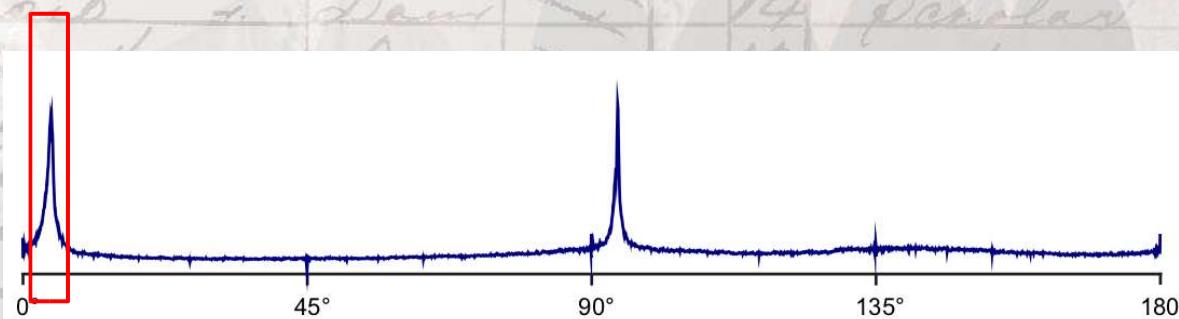
- Scale: Sample along recovered axes at exponential intervals; offset of best 1D correlation gives  $\log(\text{scale})$



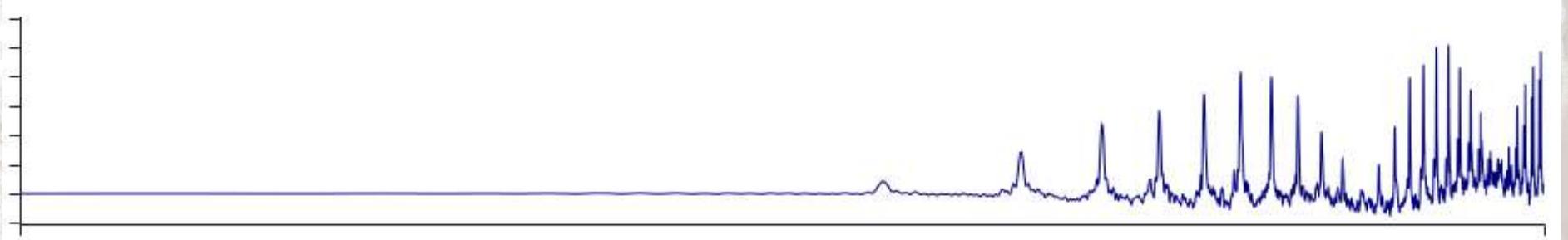
- Translation: Simple 2D correlation – needs bg removal

# Parameter Recovery

- Rotation [+ shear]: Recover document axes by finding two largest peaks in radial projection histogram



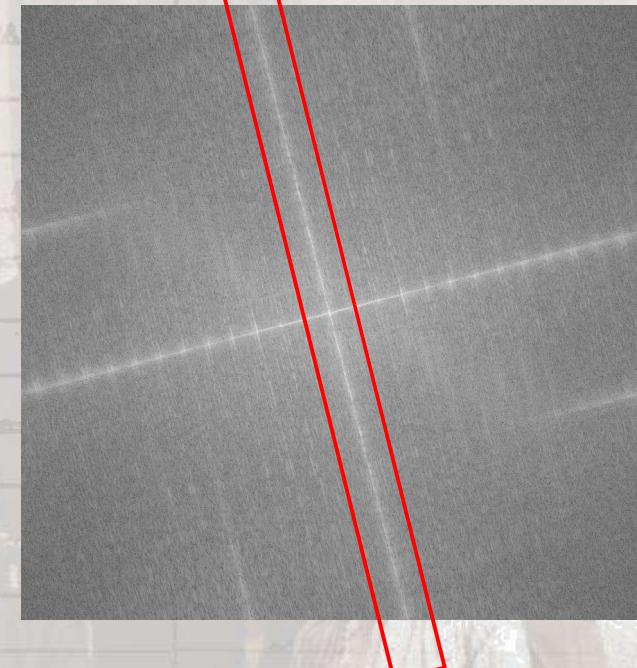
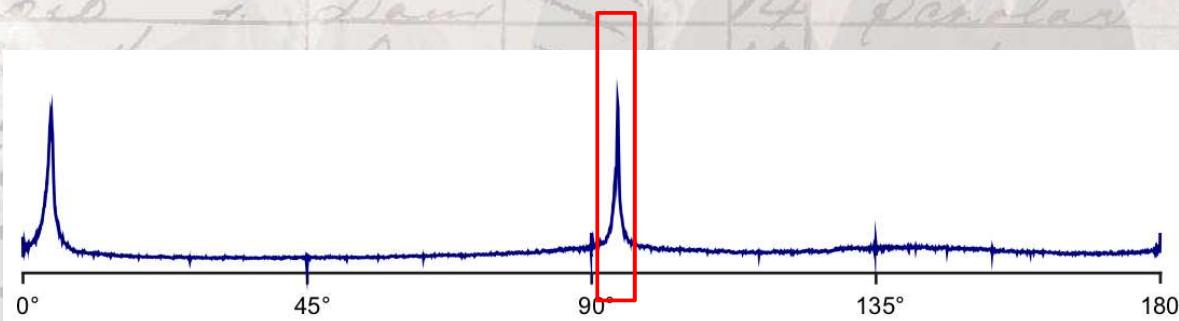
- Scale: Sample along recovered axes at exponential intervals; offset of best 1D correlation gives  $\log(\text{scale})$



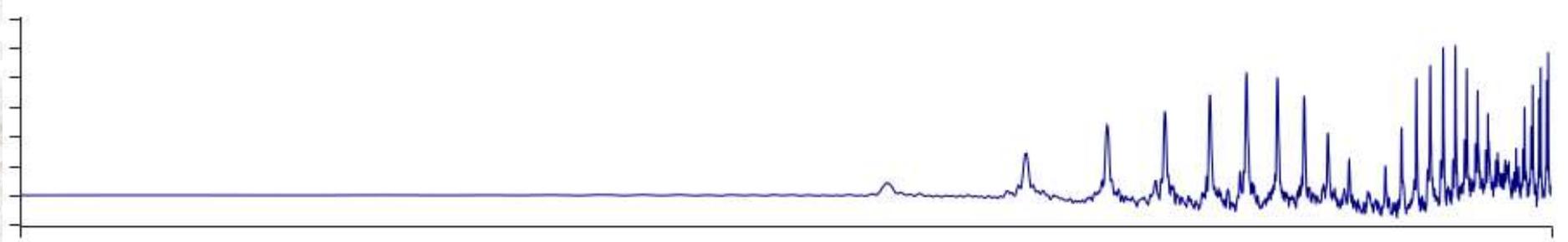
- Translation: Simple 2D correlation – needs bg removal

# Parameter Recovery

- Rotation [+ shear]: Recover document axes by finding two largest peaks in radial projection histogram



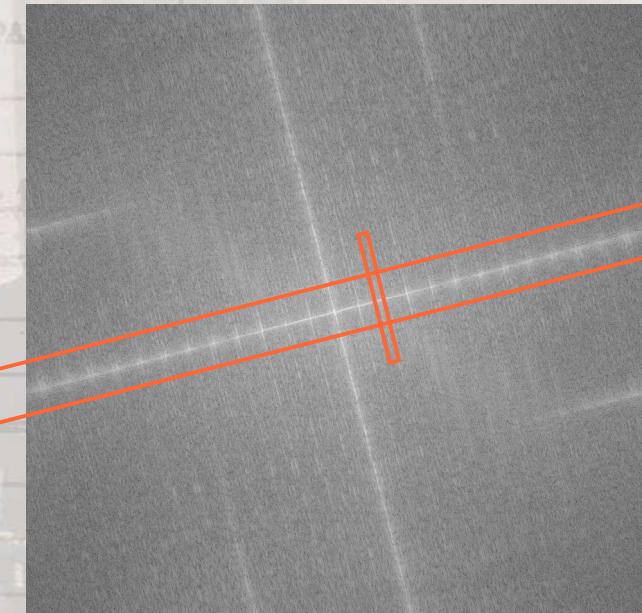
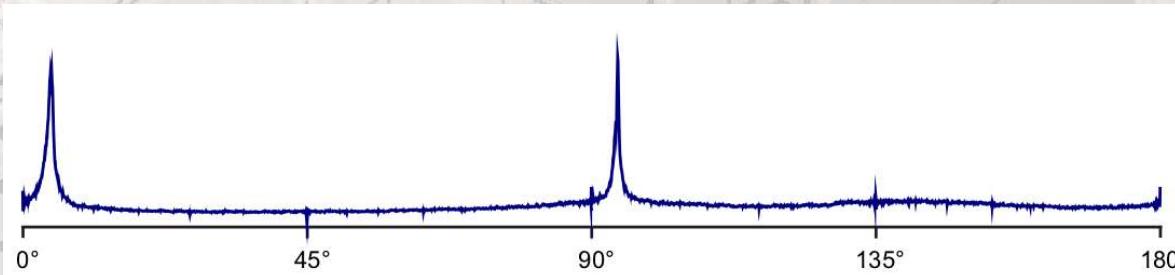
- Scale: Sample along recovered axes at exponential intervals; offset of best 1D correlation gives  $\log(\text{scale})$



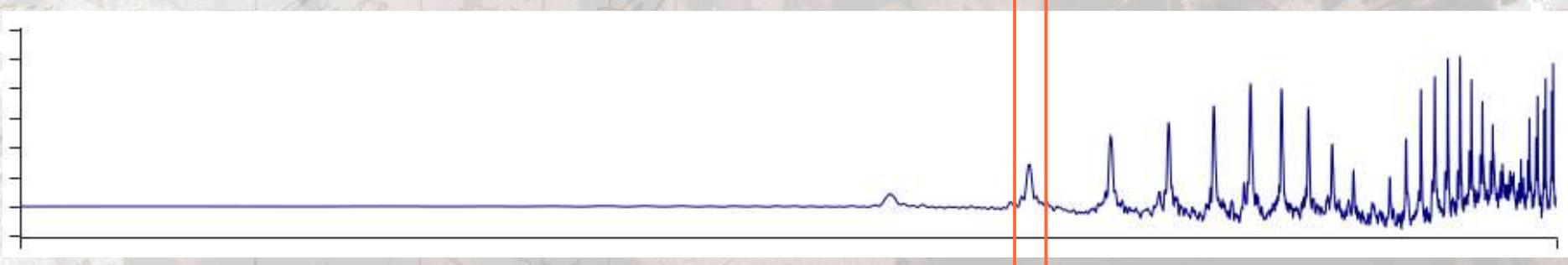
- Translation: Simple 2D correlation – needs bg removal

# Parameter Recovery

- Rotation [+ shear]: Recover document axes by finding two largest peaks in radial projection histogram



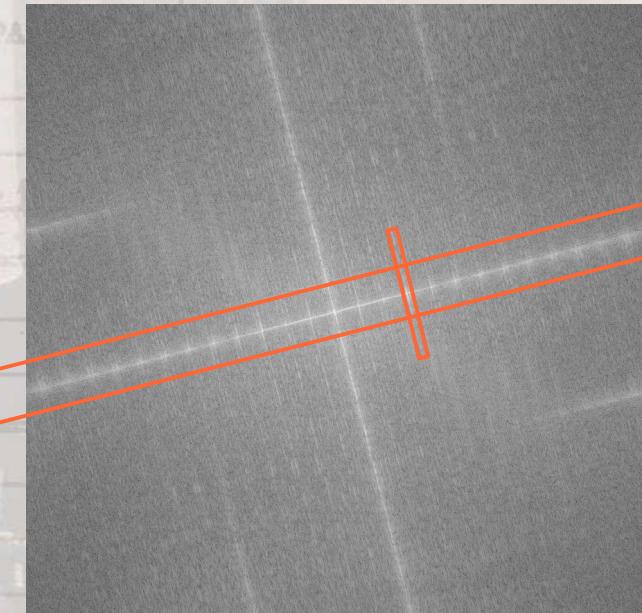
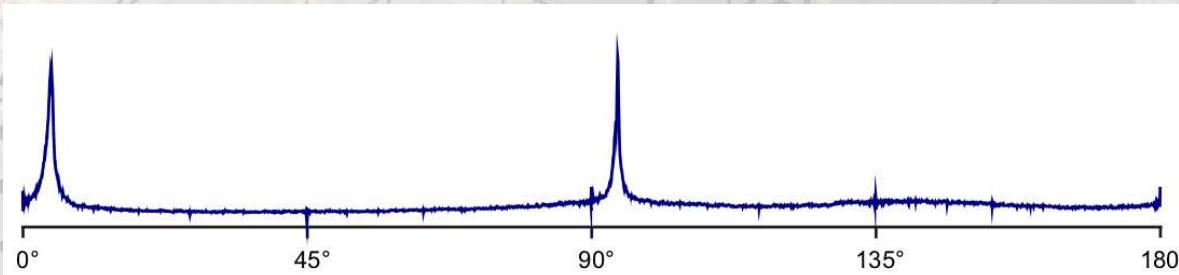
- Scale: Sample along recovered axes at exponential intervals; offset of best 1D correlation gives  $\log(\text{scale})$



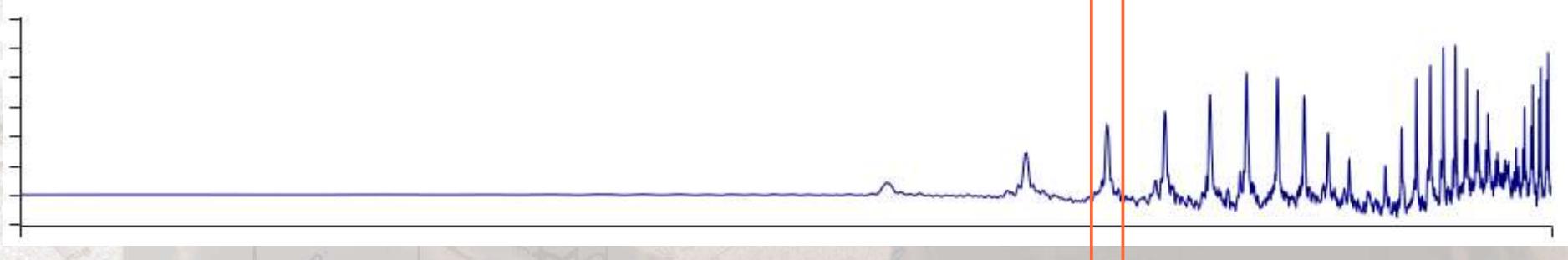
- Translation: Simple 2D correlation – needs bg removal

# Parameter Recovery

- Rotation [+ shear]: Recover document axes by finding two largest peaks in radial projection histogram



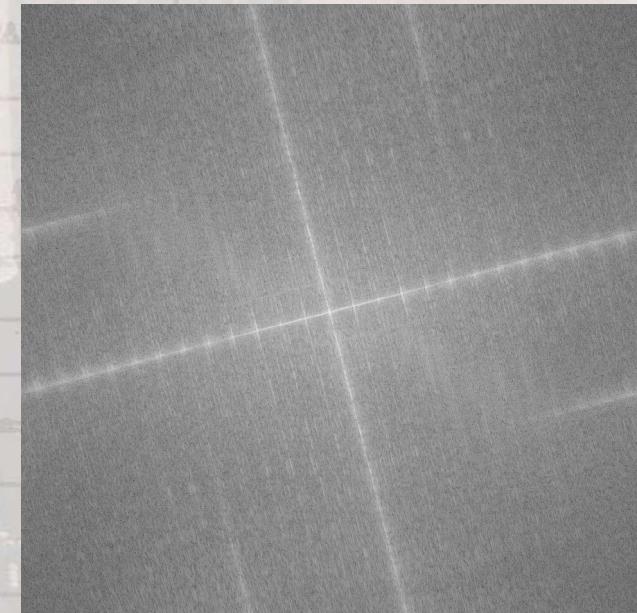
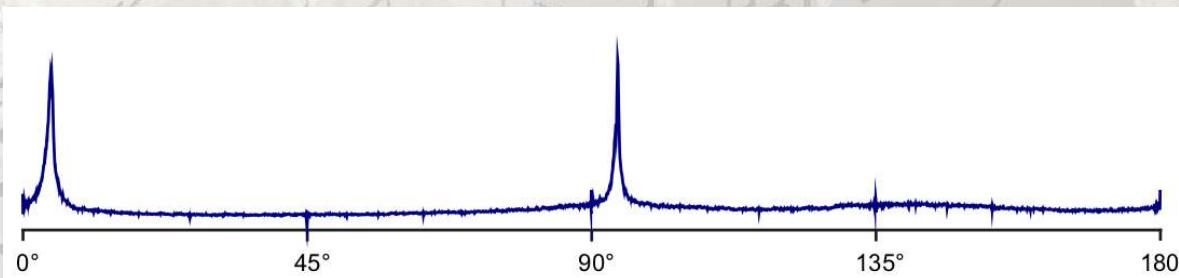
- Scale: Sample along recovered axes at exponential intervals; offset of best 1D correlation gives  $\log(\text{scale})$



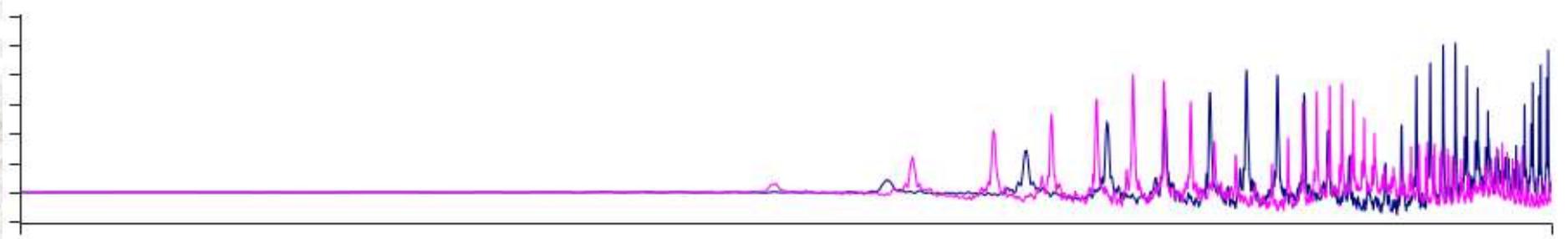
- Translation: Simple 2D correlation – needs bg removal

# Parameter Recovery

- Rotation [+ shear]: Recover document axes by finding two largest peaks in radial projection histogram



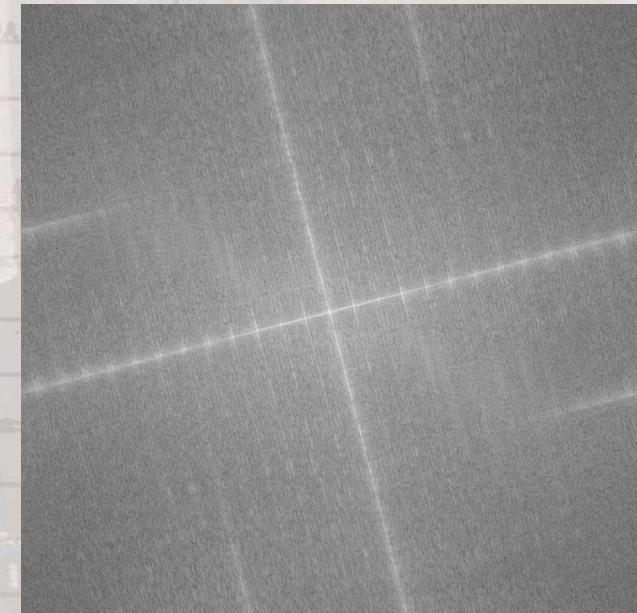
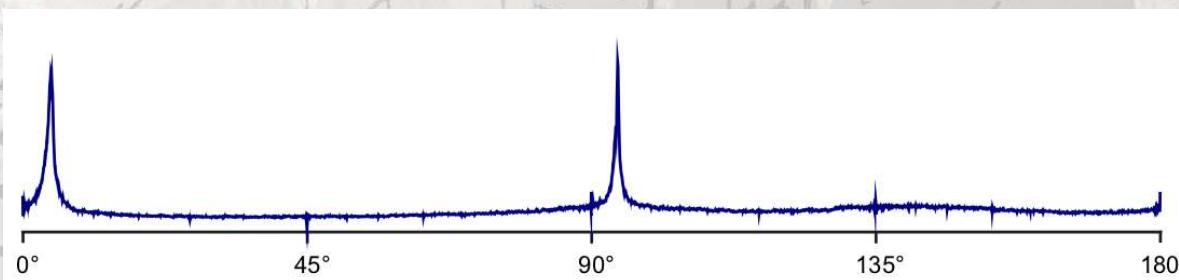
- Scale: Sample along recovered axes at exponential intervals; offset of best 1D correlation gives  $\log(\text{scale})$



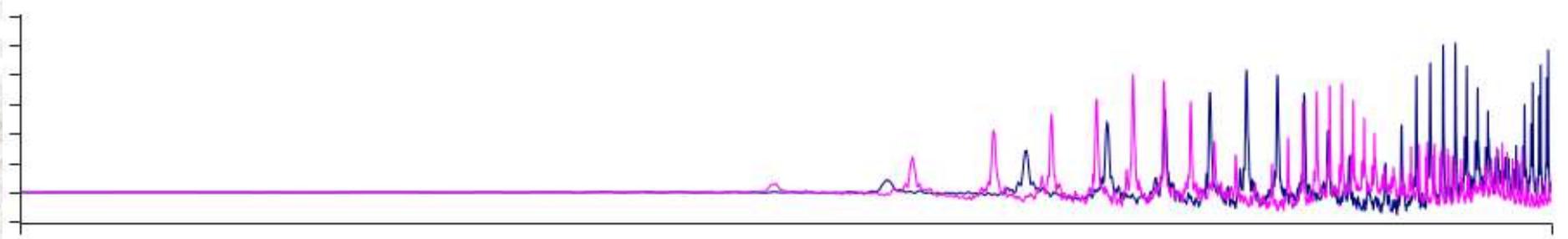
- Translation: Simple 2D correlation – needs bg removal

# Parameter Recovery

- Rotation [+ shear]: Recover document axes by finding two largest peaks in radial projection histogram



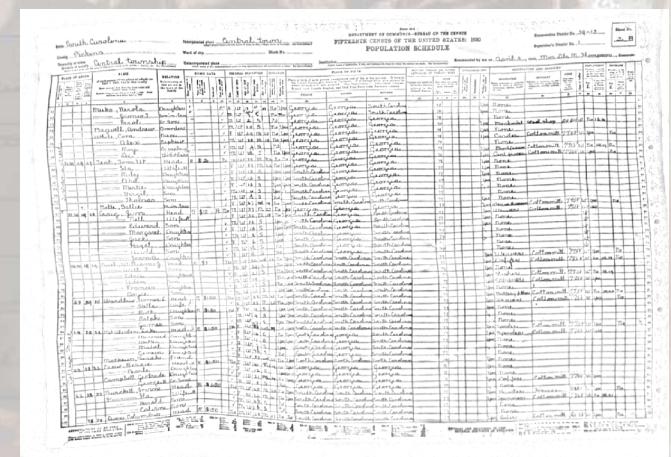
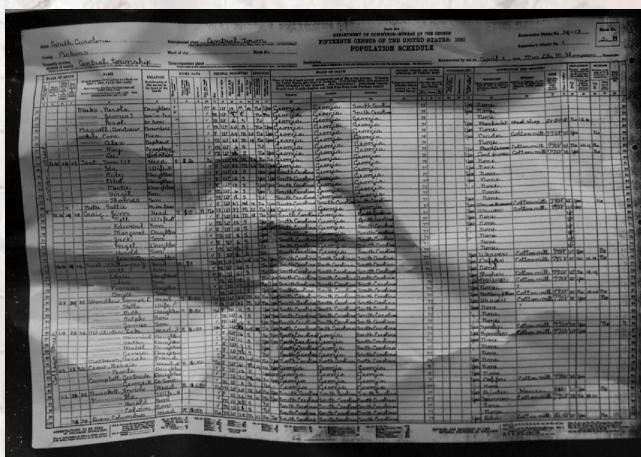
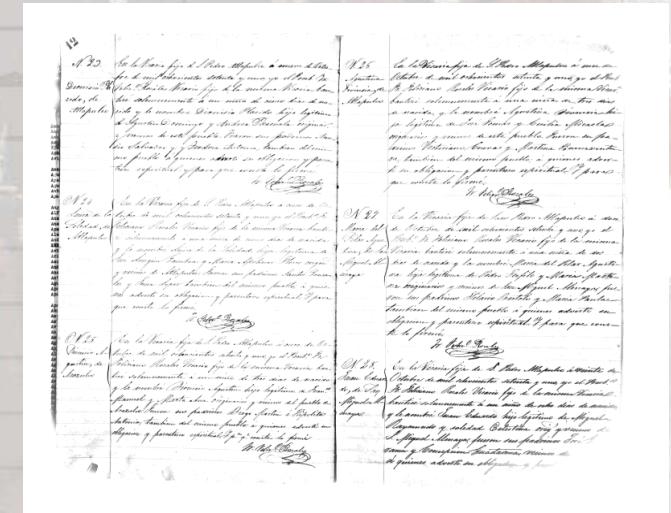
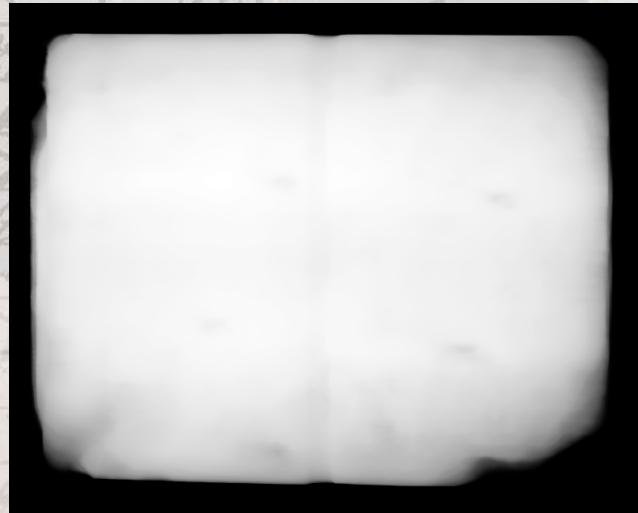
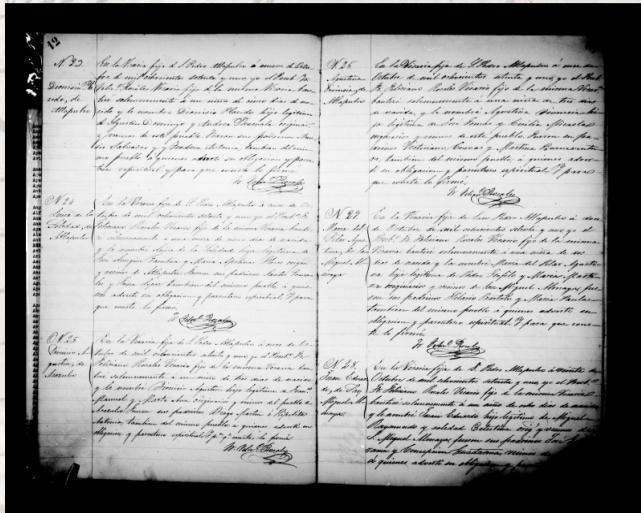
- Scale: Sample along recovered axes at exponential intervals; offset of best 1D correlation gives  $\log(\text{scale})$

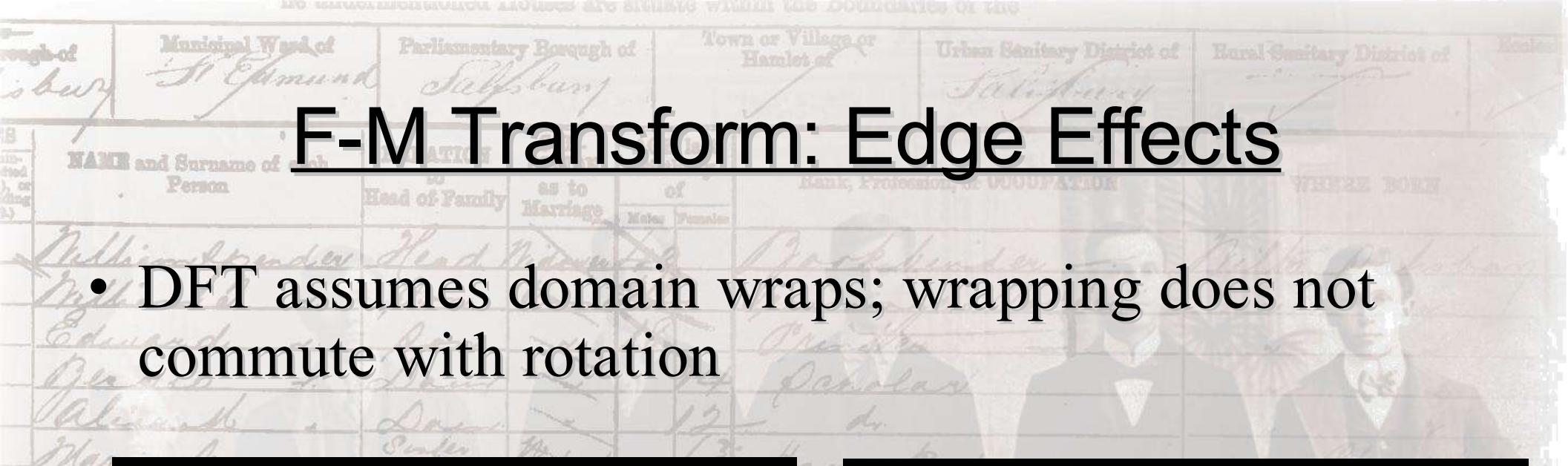


- Translation: Simple 2D correlation – needs bg removal

# Preprocessing: Automatic Background Removal

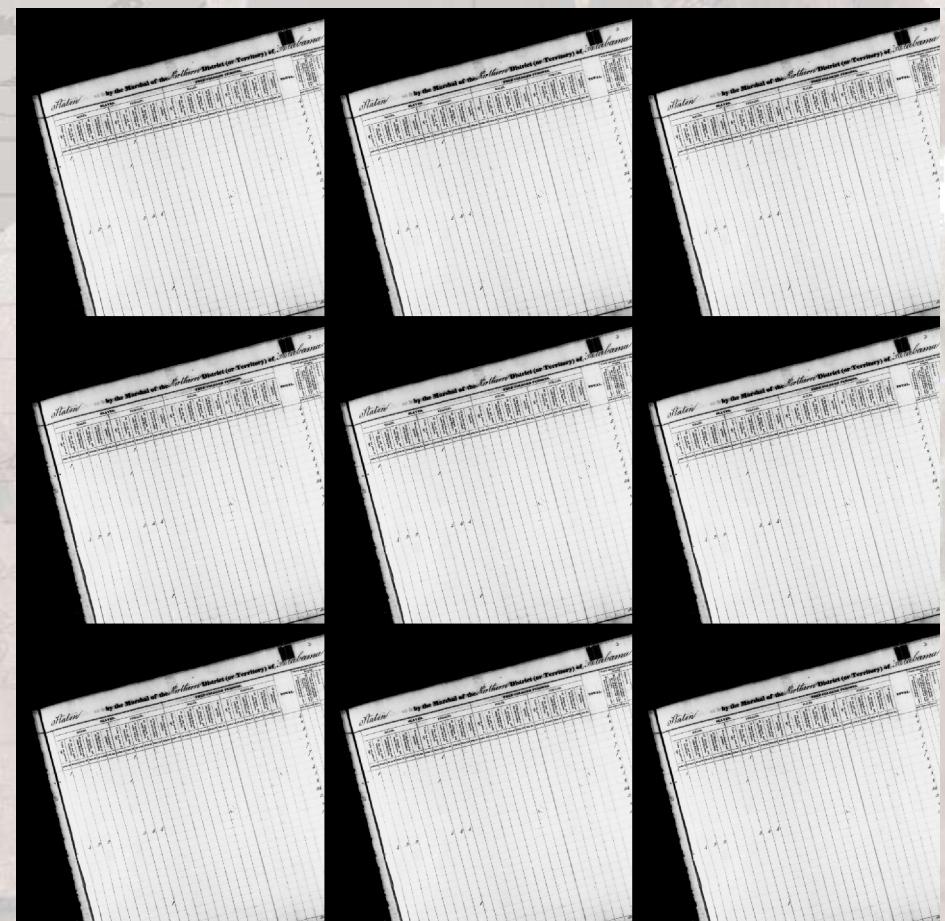
- Subtract median-filtered image from original, leaving only fine features
- Removes shading and black borders





# F-M Transform: Edge Effects

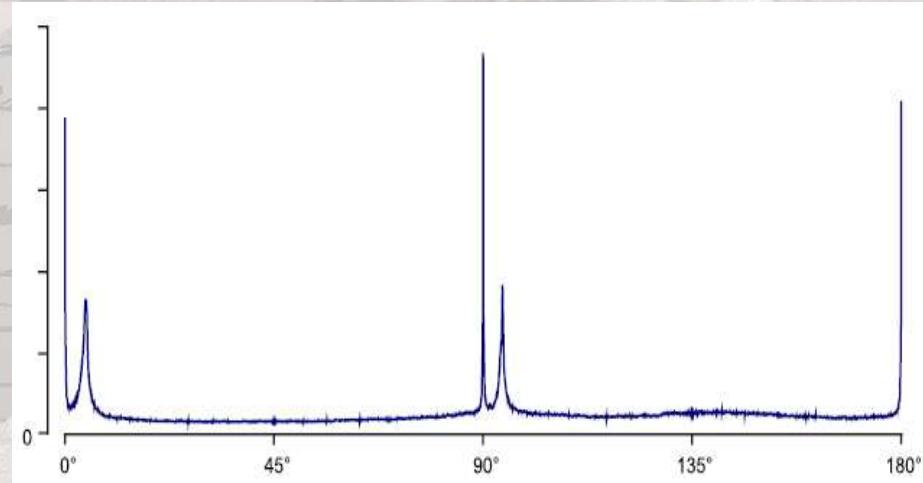
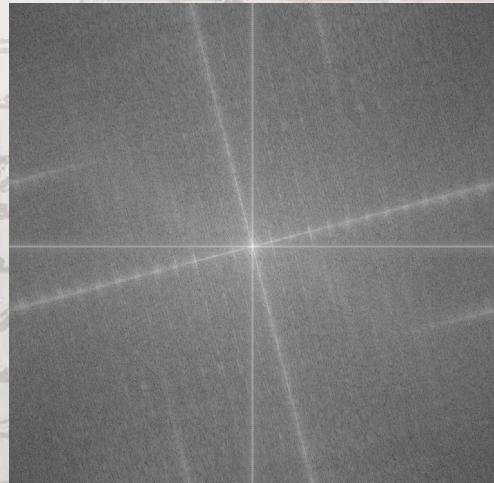
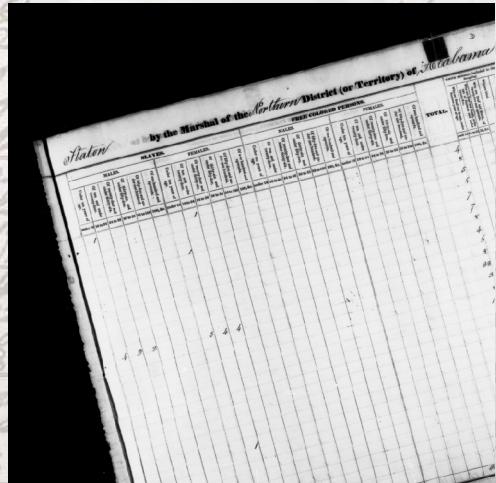
- DFT assumes domain wraps; wrapping does not commute with rotation



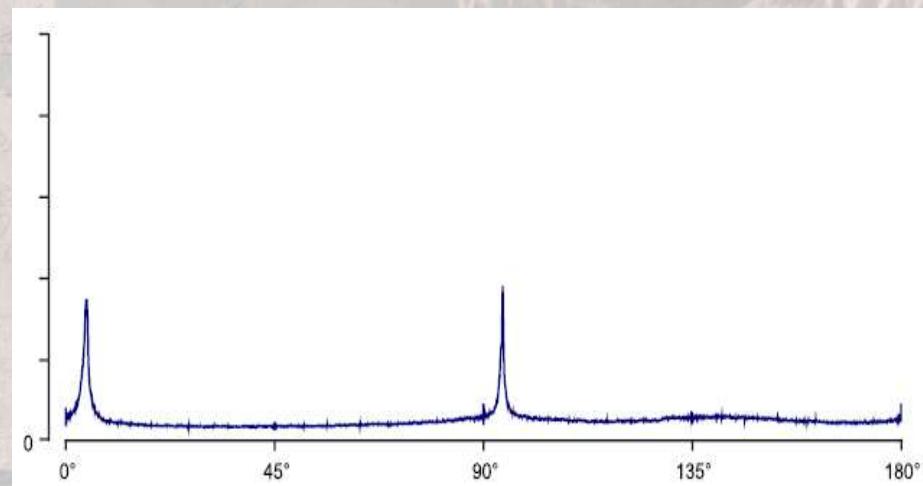
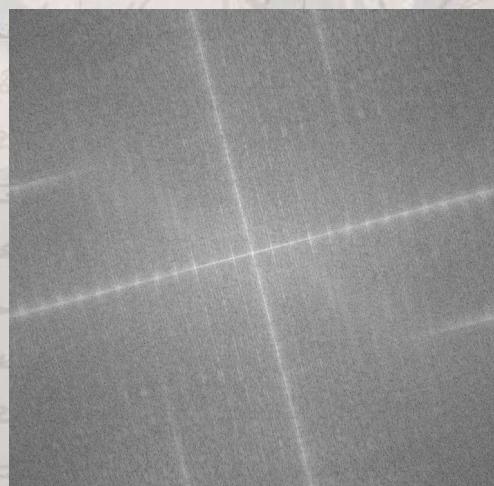
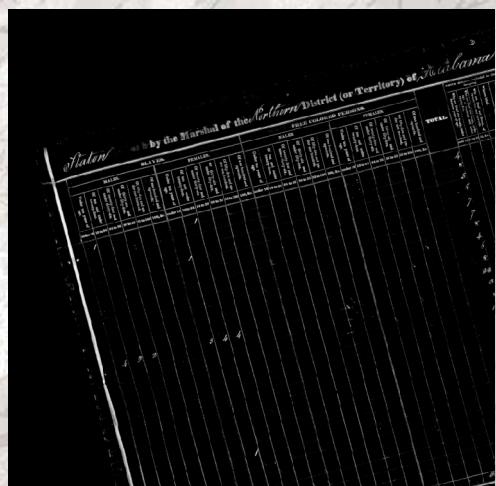
# F-M Transform: Edge Effects

Background removal eliminates the need for apodization

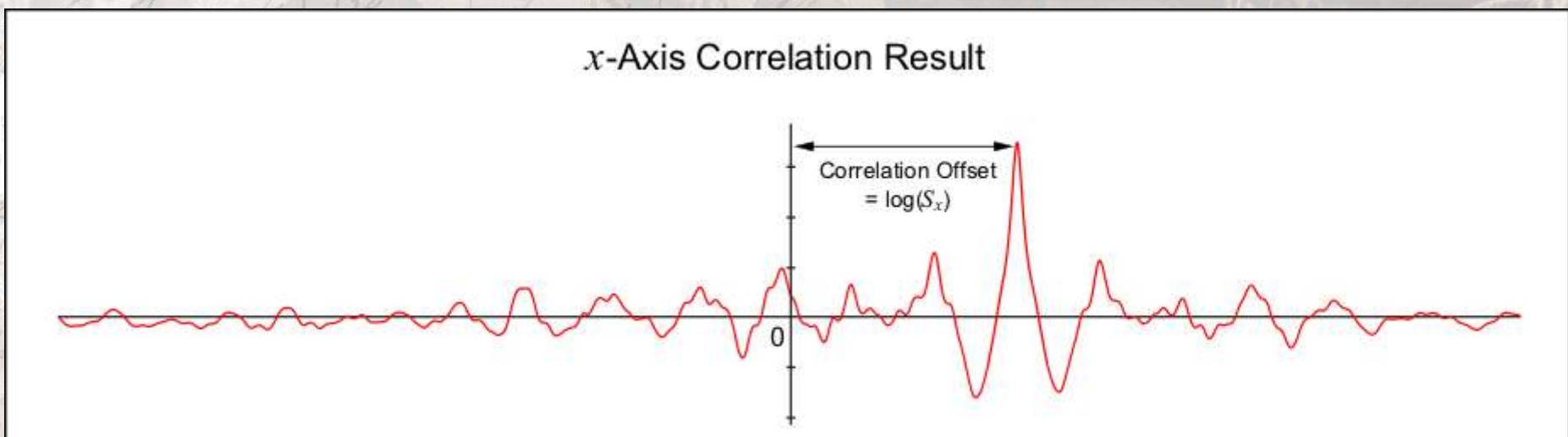
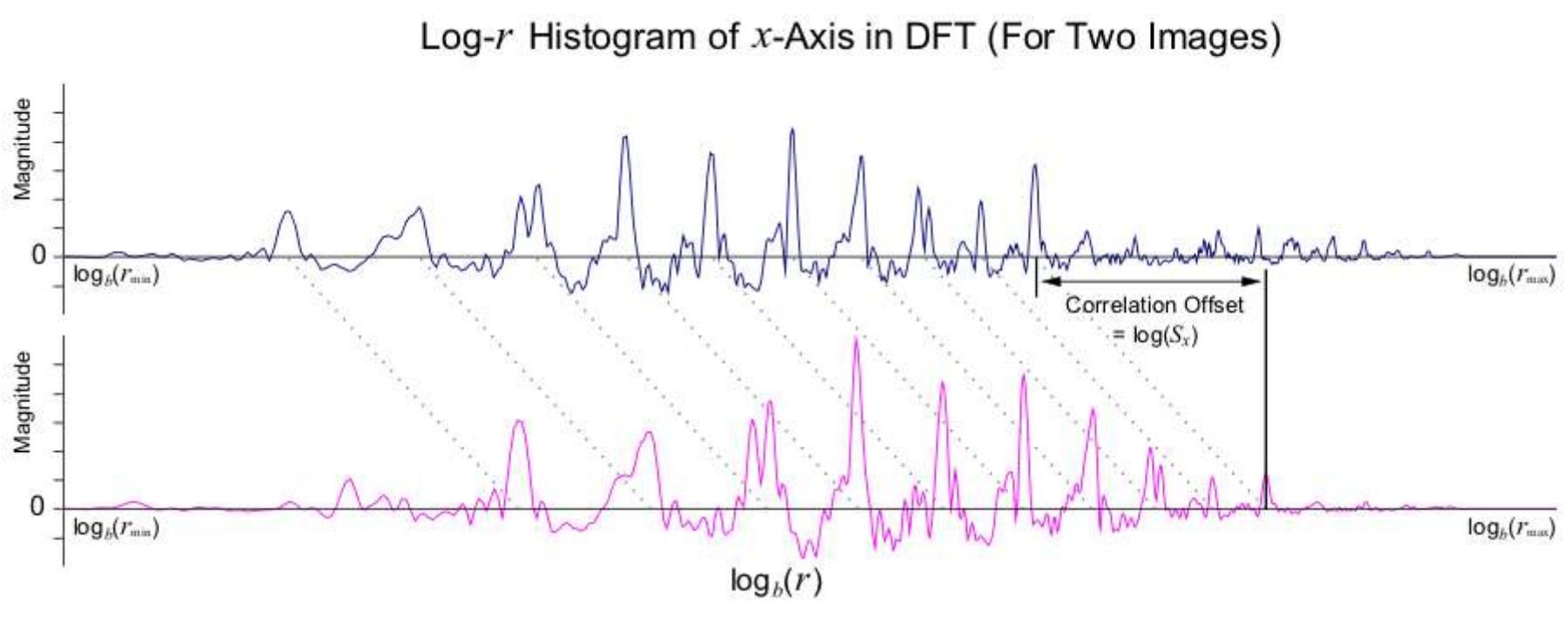
Without background removal:



With background removal:



# Scale Recovery – Example



# RMS Error

(100 Different Transformation Reversals of the Same Image)

**Rotation / Shear:**  $0.001^\circ$

**Scale:**  $0.030\%$

**Translation:**  $0.16$  pixels

## Conclusions

- Specialization of Fourier-Mellin image registration for tabular document images
- Each transform component is completely separable, and recovered in the frequency domain
- F-M registration extended to handle arbitrary affine transforms (including shear)
- Elimination of edge effects and correlation problems through median filter background removal algorithm

# Questions?

THIS IS A  
LEGAL  
RECORD AND  
WILL BE PER-  
MANENTLY  
FILED

SEE OTHER  
SIDE

FILL IN  
WITH A  
TYPEWRITER  
OR WRITE  
PRIMARILY  
WITH DARK  
INK. DO NOT  
USE GREEN  
NOR RED INK.  
LEGAL COPIES  
CANNOT BE  
MADE IF  
ENTRIES  
ARE DIM

ALL ITEMS  
MUST BE  
COMPLETE  
AND  
ACCURATE

IF NO DOCTOR  
WAS IN  
ATTENDANCE  
MEDICAL CER-  
TIFICATION  
SHOULD BE  
COMPLETED  
BY THE LOCAL  
HEALTH  
OFFICER, OR  
CORONER IF  
HE IS A  
PHYSICIAN OR  
IF INQUEST  
WAS HELD

VS-2-100M

CERTIFICATE OF DEATH  
STATE OF ALABAMA

1. PLACE OF DEATH: a. County _____		b. Street No. _____	2. USUAL RESIDENCE (Where deceased lived. If institution; residence before admission).					
			a. State _____	b. Sub-County _____				
c. City (If outside city or town limits, write RURAL) Or Town _____		d. Length of Stay (in this place) _____	d. City (If outside city or town limits, write RURAL) Or Town _____					
e. Full Name of (If not in hospital or institution, give street address Hospital Or Institution _____)		d. Street Address (If rural, give location) _____				d. Sub-County _____		
3. Name Of DECEASED (Type or Print)		a. (First) _____	b. (Middle) _____	c. (Last) _____	d. Date Of Death _____	e. Month _____	f. Day _____	g. Year _____
5. Sex _____		6. Color or Race _____	7. Married, Never Married, Widowed, Divorced (Specify) _____	8. Date of Birth _____	9. Age (In years last birthday) _____	If Under 1 Year Months _____ Days _____	If Under 24 Hrs. Hours _____ Min. _____	
10a. Usual Occupation (Give kind of work done during most of working life, even if retired) _____		10b. Kind of Business or Industry _____		11. Birthplace (State and county, or foreign country) _____	12. Citizen of What Country? _____			
13. Father's Name _____		14. Mother's Maiden Name _____						
15. Was Deceased Ever in U. S. Armed Forces? (If yes, give rank and date entered) _____		16. Social Security No. _____		17. INVESTIGATOR'S NAME AND ADDRESS _____				
18. Cause of Death Enter only one cause per line for (a), (b), and (c)		MEDICAL CERTIFICATION				Interval Between Onset and Death		
		I. Disease or Condition Directly Leading to Death* (a) _____						
		Antecedent Causes Morbid conditions, if any, giving rise to the above cause (a) stating the underlying cause last.						
		Due To (b) _____						
		Due To (c) _____						
II. Other Significant Conditions Conditions contributing to death but not related to the disease or condition causing death.								
19a. Date of Operation _____		19b. Major Findings of Operation _____		20. Autopsy? Yes [ ] No [ ]				
21a. Accident Suicide, Homicide (Specify) _____		21b. Place of Injury (Home, Farm, Factory; Street, Office, Shop, etc.) _____		21c. (City, Town, or Rural) _____ (County) _____ (State) _____				
21d. Time (Month) (Day) (Year) (Hour) of Injury _____		21e. Injury Occurred m. While at _____ Not While Work [ ] at Work [ ]		21f. How Did Injury Occur? [ ]				
22. I hereby certify that I attended the deceased from _____, 19_____, to _____, 19_____, that I last saw the deceased alive on _____, 19_____, and that death occurred at _____, from the causes and on the date stated above.								
23a. SIGNATURE (Degree: e.g.: M.D.)		23b. Address _____		23c. Date Signed _____				
24a. Burial, Cremation, Removal (Specify) _____		24b. Date _____		24c. Name of Cemetery or Crematory _____		24d. Location (City, town, or county) (State) _____		
Date Rec'd by Local Reg. _____		Registrar's Signature _____		25. Funeral Director _____		Address _____		

ALABAMA DEPARTMENT OF PUBLIC SAFETY DIVISION OF POLICE AND FIRE SERVICES

(445 images, registered)

The following returns are returnable within the boundaries of the					
Municipal Ward of St Edmund	Parliamentary Borough of Salisbury	Town or Village or Hamlet of	Urban Sanitary District of Salisbury	Rural Sanitary District of	Electoral Division
NAME and Surname of each Person	RELATION to Head of Family	CON-DITION as to Marriage	AGE last Birthday of Male Females	Bank, Profession, or OCCUPATION	WHERE BORN
William Pendleton	Head Master	Married		Bookbinder	Bethel, Wiltshire
William D.	Son	Married		Engineering Carpenter	
Edward	Son	Married		Carpenter	
Georgie	Son	Married	14	Scholar	
Salmon	Son	Married	12	Dr	
Maria Freeman	Sister	Married	63	Housekeeper	
Richard Freeman	Head Master	Married		Labourer	
James Jane	Sister	Married	39		
James Jane	Son	Married	17		
Charles	Son	Married	12	Scholar	
Alberto	Son	Married	9		
Annie	Son	Married	6		
Gerry	Son	Married	4		
John Daniels	Head Master	Married		Labourer	
John	Sister	Married			
John	Son	Married		Composition Penmanship	
John	Son	Married		General Merchant	
John	Son	Married	12	Scholar	
Matthew Pendleton	Head Master	Married	50		
John Pendleton	Son	Married	32	Engineering Carpenter	
John Pendleton	Son	Married	18		
John Pendleton	Son	Married	77		

# Pointwise Mean, Median, Percentile

NAME and Surname of each Person

RELATION to Head of Family

CON-DITION as to Marriage

AGE last Birthday of Person

Bank, Profession, or OCCUPATION

Bank, Profession, or OCCUPATION

WHERE BORN

The undermentioned Houses are situate within the Boundaries of the [Page 11]									
Civil Parish (or-Township) of	City or- Municipal Borough of	Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of	Urban Sanitary District of	Rural Sanitary District of	Religious Parish or District of		
No. of Schedule	ROAD, STREET, Acre- and No. or NAME of HOUSES	HOUSES No. of Rooms (1), or No. of Bed-Rooms (2)	NAME and Surname of each Person	RELATION to Head of Family	CON- DITION as to Marriage	AGE last Birthday of Person	Bank, Profession, or OCCUPATION	WHERE BORN	1) Dead-and-Dead 2) Married 3) Divorced or Else- where Locomotive
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									
34									
35									
36									
37									
38									
39									
40									
41									
42									
43									
44									
45									
46									
47									
48									
49									
50									
51									
52									
53									
54									
55									
56									
57									
58									
59									
60									
61									
62									
63									
64									
65									
66									
67									
68									
69									
70									
71									
72									
73									
74									
75									
76									
77									
78									
79									
80									
81									
82									
83									
84									
85									
86									
87									
88									
89									
90									
91									
92									
93									
94									
95									
96									
97									
98									
99									
100									
101									
102									
103									
104									
105									
106									
107									
108									
109									
110									
111									
112									
113									
114									
115									
116									
117									
118									
119									
120									
121									
122									
123									
124									
125									
126									
127									
128									
129									
130									
131									
132									
133									
134									
135									
136									
137									
138									
139									
140									
141									
142									
143									
144									
145									
146									
147									
148									
149									
150									
151									
152									
153									
154									
155									
156									
157									
158									
159									
160									
161									
162									
163									
164									
165									
166									
167									
168									
169									
170									
171									
172									
173									
174									
175									
176									
177									
178									
179									
180									
181									
182									
183									
184									
185									
186									
187									
188									
189									
190									
191									
192									
193									
194									
195									
196									
197									
198									
199									
200									
201									
202									
203									
204									
205									
206									
207									
208									
209									
210									
211									
212									
213									
214									
215									
216									
217									
218									
219									
220									
221									
222									
223									
224									
225									
226									

# Apodization

Attenuation to zero at the edge of the image, to reduce sharp transitions  
on wrapping

