

Zoning Tabular Documents

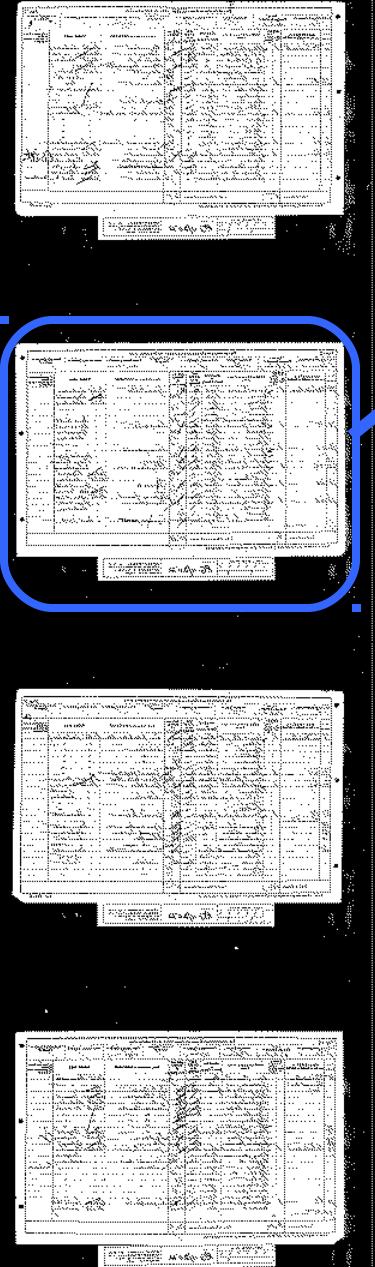
Heath Nielson and William Barrett

Motivation

- Move granularity of indexing from image level to field level
 - Search or browse through fields rather than images
- Let the computer perform the repetitive task of finding regions within a document and determining content of those regions

Processing Pipeline

Cropping



[Example]

The undermentioned Houses are situate within the Boundaries of the
Parliamentary Borough of
R. May

18

No. of Houses	ROAD STREET, Ac- cording to the Name of HOUSE	NAME and Surname of each Person	RELATION as to Head of Family	AGE last on 1st JUNE of the Year	SEX Males Females	Rank, Profession, or OCCUPATION	WHERE BORN	Civil Parish or Town or Village or Hamlet of R. May		Municipal Ward of Wales	Parliamentary Borough of R. May	Town or Village or Hamlet of R. May	Urban Sanitary District of R. May	Rural Sanitary District of R. May	Ecclesiastical Parish or District of R. May
								1	2	3	4	5	6		
1	7 Charlotte St. ("Queen's Arms")	Michael Morris	Brood	47	Mrs.	Licensed Victualler	Middlesex; Paddington								
		Mary J. Morris	Wife	39			Suffolk; Colchester								
		Eliza	Daughter	7			Suffolk; Newmarket								
		Elizabeth Morrison	Daughter	58		Assistant	Suffolk; Ipswich								
		Ann Fox	Daughter	28		General Serv.	Herts; Stevenage								
		Catherine Doyle	Daughter	28		Barmaid	Herts; Stevenage								
2	8 Charlotte St.	Donald Newton	Brood	37	Mrs.	Glass; (makes, employing 2 men)	Cambridgeshire; Weyton								
		Eliza	Wife	36			Essex; Longfawne								
		William	Son	12		Soldier	Suffolk; Sudbury								
		Beatrice	Daughter	1		Dr.	Essex; Drayton								
		George Brown	Servant	19		General Serv.	Hertfordshire; Hertford								
		James Cook	Daughter	22		General Serv.	Suffolk; Sudbury								
3		James F. Hollings	Brood	41		Baker's Clerk	Yorks; Leeds								
		Harriet	Wife	39			Essex; Braintree								
		Stephen White	Daughter	18		General Serv.	Suffolk; Bideford								
4	7, Broad Lane	William Pennington	Brood	72	Mrs.	Cook Farmer	Hertfordshire; Hertford								
		Anna	Wife	69			Essex; Tiverton								
5	2, Broad Lane	Thomas Johnson	Brood	65		Porter	General Serv.								
		Henry Johnson	Daughter	39			Essex; Stevenage								
		Emma	Daughter	18		Coat Maker	Suffolk; Sudbury								
		Jane Farmer	Servant	18		Coat Maker (Apprentice)	Herts; St. Albans								
		Walter Campbell	Lodger	23		Ship Carpenter (out of employ)	Suffolk; Sudbury								
						End of St. M. List	Cambridgeshire District								
Total of Houses...	146	Total of Males and Females...	938												

PUBLIC RECORD OFFICE

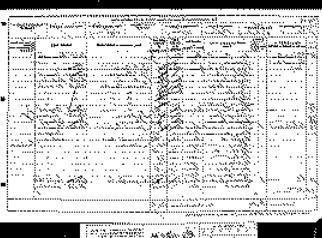
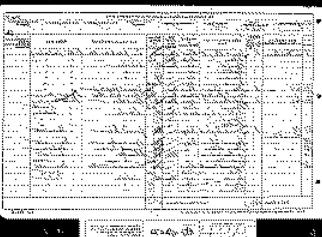
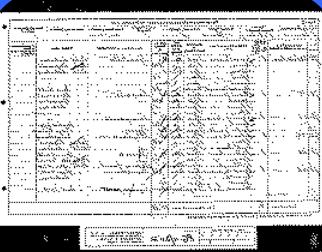
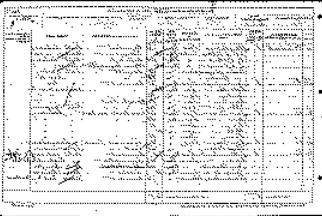
1	2	3	4	5	6
7	8	9	10	11	12

Reference No.: RG 11/2072

COPYRIGHT PHOTOGRAPH - NOT TO BE
REPRODUCED PHOTOGRAPHICALLY WITHOUT
PERMISSION OF THE PUBLIC
RECORD OFFICE, LONDON

Processing Pipeline

Zoning



[Example]

The undermentioned Houses are situate within the Boundaries of the
Civil Parish or Township of R. May.

18

No. of Houses	ROAD STREET, Ac. and No. or NAME of HOUSE	NAME	RELATION of each Person to Head of Family	AGE last Year of each Person	Rank, Profession, or OCCUPATION	WHERE BORN	Ecclesiastical Parish or District of R. May.	
							St. Michael's - Drury St. Peter's - Drury St. Paul's - Drury St. John's - Drury	St. James' - Drury St. Peter's - Drury St. Paul's - Drury St. John's - Drury
1	7, Charlotte St. ("Queen's Arms")	Michael Morrison	Head	37	Licensed Victualler	Middlesex; Paddington		
		Mary Do.	Wife	32		Servt; Confectioner		
		Ellen Do.	Daughter	7		Servt; Shoemaking		
		Elizabeth Do.	Daughter	5	Assistant	Servt; Shoemaking		
		Ann Do.	Sister	2	General Serv.	Herts; Teddington		
		Catharine Doyle	Sister	2	General Serv.	Herts; Teddington		
		Elizabeth Weston	Sister	37	General Serv.; (maiden, employing 2 men)	Camden; Weyton		
		James Do.	Wife	36		Do.; Longsaw		
		William Do.	Son	12	Soldier	Servt; Teddington		
		Elizabeth Do.	Sister	1		Do.; Do.		
		George Brown	Servt	19	General Servant	Middlesex; Islington		
		Jane Do.	Servt	22		Servt		
		James F. Hollings	Servt	41	Baker's Clerk	Yorkshire; Leeds		
		Harriet Do.	Wife	39		Do.; Shoefit		
		Stephen White	Servt	18	General Serv.	Servt; St. Pancras		
		John Do.	Head	72	Cook Farmer	Hertfordshire; Hertford		
	7, Broad Lane	William Pennington	Wife	69		Do.; Townsmith		
		Anna Do.	Daughter	69				
	8, Broad Lane	Thomas Johnson	Head	65	Porter; Grocer	Leicestershire; Kesteven		
		Henry Johnson	Son	39		Servt; Shoemaking		
		Emma Do.	Servt	18	Cook	Middlesex; St. Pancras		
		Jane Foster	Servt	25	Cook; Baker (Apprentice)	Servt; Dublin		
		Walter Campbell	Lodger	25	Ship Carpenter; (not of employ)	Durham; Sunderland		
	Total of Houses...	146			End of St. Michael's			
		238			Continuation District			
					Total of Males and Females... 912			

PUBLIC RECORD OFFICE

RG 11/2072

Reference:

COPYRIGHT PHOTOGRAPH - NOT TO BE
REPRODUCED PHOTOGRAPHICALLY WITHOUT
PERMISSION OF THE PUBLIC
RECORD OFFICE, LONDON

NAME and Surname of each Person

Michael Morrison

Mary J. Do.

Ellen Do.

Processing Pipeline

Recognition

[Example]

The undermentioned Houses are situate within the Boundaries of the
Civil Parish or Township of R. May
in the County of Lancashire.

18

No. of Households	ROAD STREET, AC. AND NO. OR NAME OF HOUSE	NAME & SURNAME OF EACH PERSON	RELATION OF EACH PERSON TO HEAD OF FAMILY	AGE last BIRTHDAY of Person	RANK, PROFESSION, OR OCCUPATION	WHERE BORN
1	7 Charlotte St. ("Queen's Arms")	Michael Morrison	Head	37	Licensed Victualler	Middleton, Lancaster
		Mary J. Do.	Wife	32	Soldier, Guards	Soldier, Lancaster
		Ellen Do.	Daughter	7	Soldier, Lancaster	Soldier, Lancaster
		Elizabeth Do.	Sister	58	Assistant	Soldier, Lancaster
		Ann Do.	Sister	28	General Serv.	Hornby, Lancaster
		Catherine Doyle	Sister	28	General Serv.	Hornby, Lancaster
		Thomas Doyle	Sister	28	General Serv.	Hornby, Lancaster
		William Doyle	Sister	12	Soldier	Soldier, Lancaster
		Katherine Doyle	Sister	1		
		George Brown	Servant	19	General Serv.	Middleton, Lancaster
		James Cook	Servant	22	General Serv.	Soldier
		James F. Hollings	Head	41	Banker's Clerk	Gateshead, Leeds
		Harriet Do.	Wife	39	Servant	Do., Bradford
		Stephen White	Sister	18	General Serv.	Soldier, Bradgate
		John Doyle	Head	72	Cook, Farmer	Houghall, Huddersfield
		William Pennington	Wife	69		Do., Penwortham
		Anna Do.	Daughter	65	Postman, Grocer	Gateshead, Huddersfield
		Thomas Johnson	Son	39	Coygher	Soldier, Lancaster
		Henry Johnson	Son	38	Cook, Baker	Middleton, St. Georges
		Emma Do.	Sister	38	Cook, Baker	Middleton, St. Georges
		John Farmer	Servant	28	Cook, Baker (Apprentice)	Soldier, Lancaster
		Walter Campbell	Servant	23	Ship Carpenter (not of employ)	Darland, Sunderland
	Total of Houses...	146	Total of Males and Females...	328		

PUBLIC RECORD OFFICE
Reference: RG 11/2072

COPYRIGHT PHOTOGRAPH - NOT TO BE
REPRODUCED PHOTOGRAPHICALLY WITHOUT
PERMISSION OF THE PUBLIC
RECORD OFFICE, LONDON

NAME and Surname of each Person

Michael Morrison

Mary J. Do.

Ellen Do.



NAME and Surname of each Person

Michael Morrison

Mary J.

Ellen

Zoning Tabular Documents

No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES In- habit- ed (U.) or Building (B.)	NAME and Surname of each Person	RELATION to Head of Family	CON- DITION as to Marriage	AGE last Birthday of Males Females	Rank, Profession, or OCCUPATION	WHERE BORN	If (1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic
28	15 Church St.	1	Ann Grant	Head	W		50	Anniutant	Salisbury
			Margaret Grant	Dau	Unm		22	Domestic Servant	Kent
			Annie Grant	Dau	Unm		18	do. do.	Salisbury
			Austin Grant	Son	Unm	20		Hawker	Kent
29	72 Salt Lane	1	Emily Morgan	Head	W		34	Shopkeeper	Norfolk, Norwich
			Alfred Morgan	Son	Umn	12		Scholar	Andover
			Hubert Morgan	Son	Umn	9		do.	do.
			Walter Morgan	Son	Umn	7		do.	do.
			Emily Morgan	Dau	Umn		4	do.	do.

Profiles

- Horizontal profile $p_h(y) = \sum_{i=0}^M \text{image}(i, y)$
- Vertical profile $p_v(x) = \sum_{i=0}^N \text{image}(x, i)$

The undermentioned Houses are situate within the Boundaries of the

Civil Parish or Township St. Mary	City or Municipal Borough of Bromley	Municipal Ward of Wells	Parliamentary Borough of Bromley	Town or Village or Hamlet of Bromley	Urban Sanitary District of Bromley	Rural Sanitary District of Bromley	Reckoned Parish District of St. Michael
No. of Schedule	HOUSES No. and Name of HOUSE	NAME and Surname of each Person	RELATION to Head of Family	GEN- DER as to Marriage	AGE last Birthday from January	Rank, Profession, or OCCUPATION	WHERE BORN
3	J. Charlotte St. 10 Anna's Arms T.	Richard Morris Mary J. Dr. Ellie Dr. Elizabeth Morris Ann Fox Catherine Doyle	Head Wife Daughter Mother Son Son	Mr. Mrs. Miss W. Son Son	34 39 7 55 28 26	Dividend Purchaser Sister, Cook Sister Assistant General Serv. Barmaid	Bromley; Folkestone Sister, Cook Sister, Browning Sister, Browning Folkestone Folkestone Folkestone
5	S. Charlotte St.	Louisa Foster Emma Dr. William Dr. Frances Dr. George Brown Jane Cook James F. Phillips Kirstie Dr. Lydia White	Head Wife Son Daughter Daughter Son Son Son Son	Mrs. Mrs. Dr. Mrs. Wife Mrs. Dr.	39 36 12 9 19 22 51 29 16	General Serv. (master, employing 2 men) Cook Folkestone Folkestone Grocer's Assistant General Serv. Book-keeper Sister General Serv.	Canterbury; Weymouth Dr.; Langdon Sister, Folkestone Dr.; Dr. Maid-servant Folkestone Folkestone Grocer's Assistant Folkestone Folkestone Folkestone
7	1, Blue Lane	William Thompson Jane Dr. Thomas Johnson Henry Johnson Emma Dr. Jane Foster Walter Campbell	Head Wife Head Son Son Apprentice Lodger	Mr. Mrs. Wife Son Son Wife Wife	72 69 68 59 51 18 23	Cashier, Tailor Sister Folkestone Folkestone Cashier Cashier (Apprentice) Help Carpenter (out of employ)	Suffolk; Bishop Dr.; Sonnenthal Dover; Kentish Sister, Bromley Maid-servant; St. Pancras Folkestone Berkhamsted
8	2, Blue Lane						
9	3a						
Total of Houses.		4	18	Total of Males and Females.	9	11	

PUBLIC RECORD OFFICE						Reference-
1	2	3	4	5	6	RG 11/2072.
1	1	1	1	1	1	

COPYRIGHT PHOTOGRAPH - NOT TO BE
REPRODUCED PHOTOGRAPHICALLY WITHOUT
PERMISSION OF THE PUBLIC
RECORD OFFICE, LONDON

Matched Filter

Creation

- Get 3 samples from the profile containing the highest “peak”
- Determine the number of points on either side of the “peak” to establish the size of the filter
- Set the value at each point in the filter to the average value from the corresponding points in the 3 samples
- Compute the average value from each of the filter’s points and subtract that amount from each point in the filter

The undermentioned Houses are situate within the Boundaries of the

Civil Parish (or Township) of St. Mary	City or Municipal Borough of Bury	Municipal Ward of Bury	Petropolitan Borough of Bury	Town or Village or Hamlet	Urban Sanitary District of Bury	Rural Sanitary District	Religious Parish or District of St. Michael
No. & House	ROAD, STREET, & NUMBER or NAME of HOUSE	NAME and Surname of each Person	RELATION to Head of Family	AGE in YEARS as on March 1871	SEX and MARRIAGE	Trade, Profession, or OCCUPATION	WHEEL HORSE
5	J. Charlotte St.	Richard Morris	Head	Mr.	55	Clerical Officer	Middleton; Edgerton
	(John's Sons)	Henry J. Mr.	Wife	Mr.	39		Male; Coalman
		John Dr.	Son	Mr.	21		Male; Shoemaking
		Elizabeth Morris	Sister	Mr.	27	Domestic	Male; Drawing
		John Fox	Son	Mr.	25	General Serv.	Male; Labourer
		Catherine Leigh	Son	Mr.	24	General	Male; Fisherman
5	E. Charlotte St.	Richard Morris	Head	Mr.	59	General; employing 2 men	Connelton; Wigan
		Eliza Dr.	Wife	Mr.	36		Dr.; Carpenter
		William Dr.	Son	Mr.	22	Labourer	Male; Labourer
		Elizabeth Dr.	Son	Mr.	21	Labourer	Male; Labourer
		George Morris	Sister	Mr.	19	General Serv.	Middleton; Skilled
		John Cook	Son	Mr.	23	General Serv.	Male; Labourer
6	James St. Whalley	Richard Morris	Head	Mr.	41	Banker's Clerk	Whalley; Bank
		Eliza Dr.	Wife	Mr.	39		Dr.; Banker
		Lydia White	Son	Mr.	18	General Serv.	Male; Bridgeman
7	St. Paul's	William Thompson	Head	Mr.	39	General Servicer	St. Paul's; Worker
		Anna Dr.	Wife	Mr.	36		Dr.; Domestic
8	St. Paul's	Thomas Johnson	Head	Mr.	65	Farmer; Game	St. Paul's; Farmer
		Henry Johnson	Son	Mr.	39	General	Male; Skilling
		James Dr.	Son	Mr.	27	General Worker	Middleton; St. Paul's
		John Farmer	Sister	Mr.	47	General Worker (affectionate)	Male; Labourer
		Walter Campbell	Sister	Mr.	22	High Carpenter (and of employ)	St. Paul's; Skilled
		Total of St. Michael Ecclesiastical District					
	Total of Houses.	4	100				
	Total of Males and Females.	35	19				

PUBLIC RECORD OFFICE					
1	2	3	4	5	6

Reference

RG 11/2072

OFFICIAL PHOTOGRAPH - NOT TO BE
REPRODUCED PHOTOGRAPHICALLY WITHOUT
PERMISSION OF THE PUBLIC
RECORD OFFICE, LONDON

Geometric Layout

- Split the document into its component parts, representing similar geometric layouts:
 - Header
 - Body
 - Footer

Body Identification

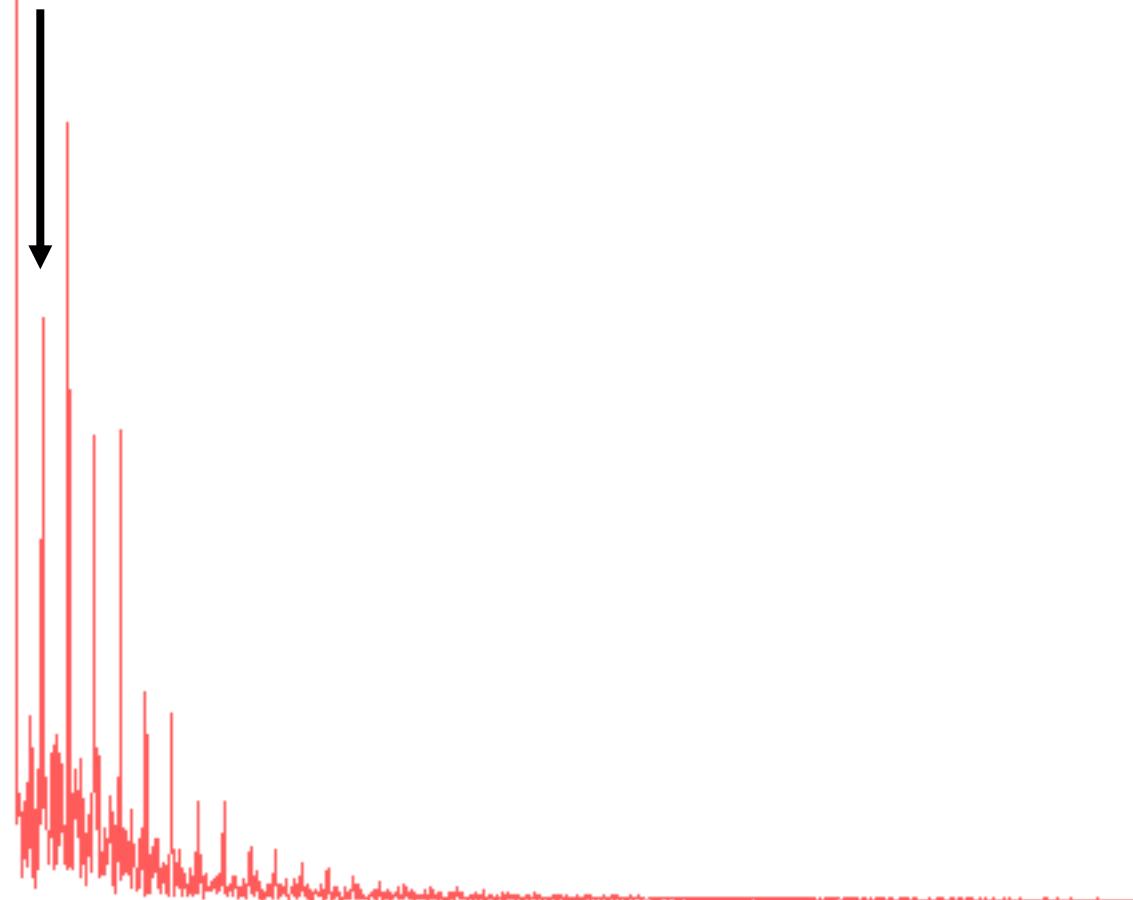
- Exploit the periodicity of the rows
- Compute $P_h(s) = \Im(p_h(y))$
- Identify first peak (lowest frequency)
- Compute $w = ps / f$
- Identify lines using the 2-prong probe

Body Identification



Horizontal Profile

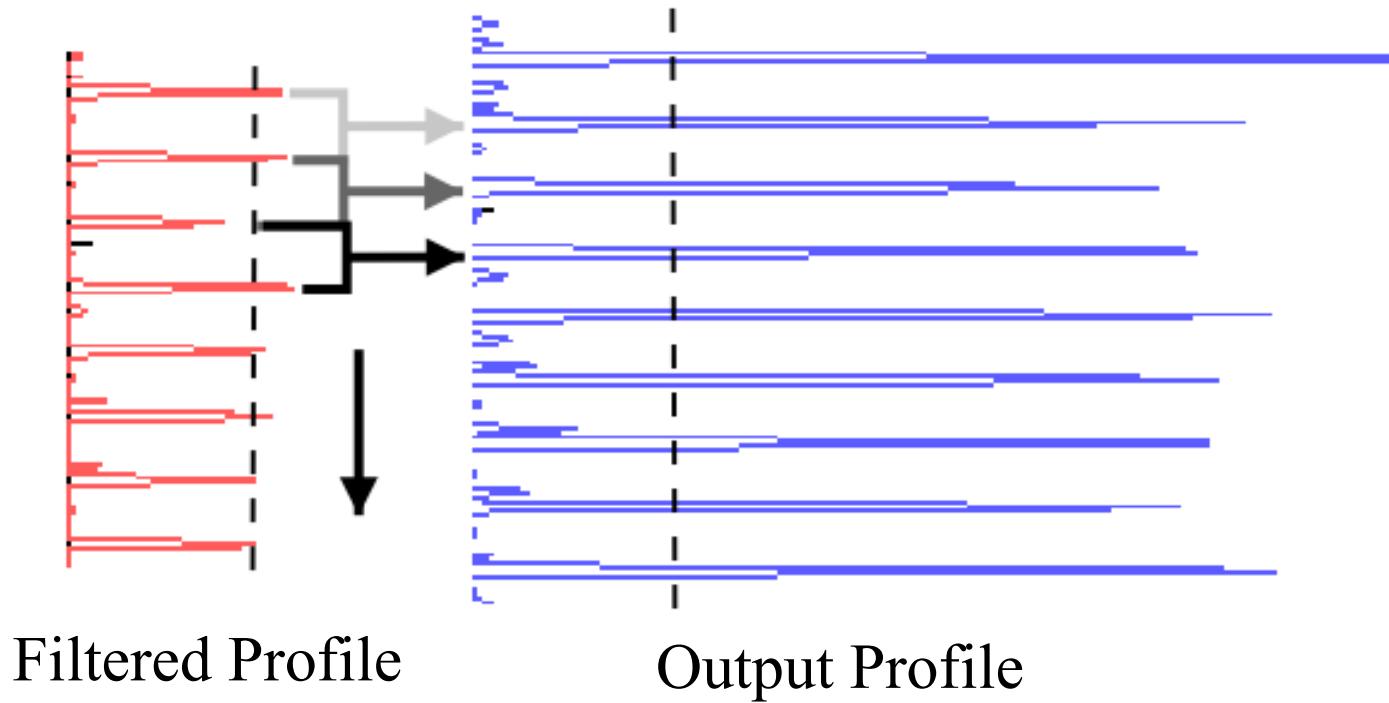
Lowest peak frequency



Amplitude Spectrum

Body Identification

2-Prong Probe



$$C(i) = \sum_{j=i-\delta}^{i+\delta} p_h(j) + \sum_{j=i-\delta}^{i+\delta} p_h(j-w)$$

Body Line Classification

Intra-document Consensus

1	1. George R?	1	William White	Head	M
			Martha White	Wife	M
			Edward White	Son	
			Annie White	Daughter	
			Arthur White	Son	
			John White	Daughter	

Row Candidates

Green row identified as false positive

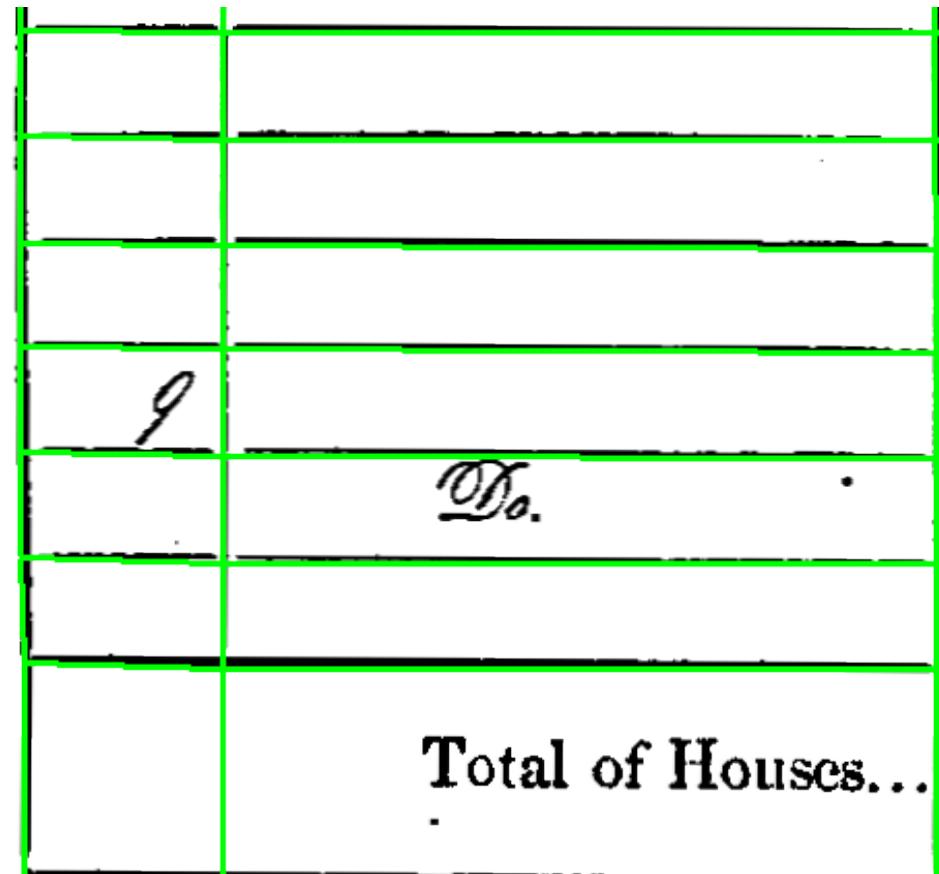
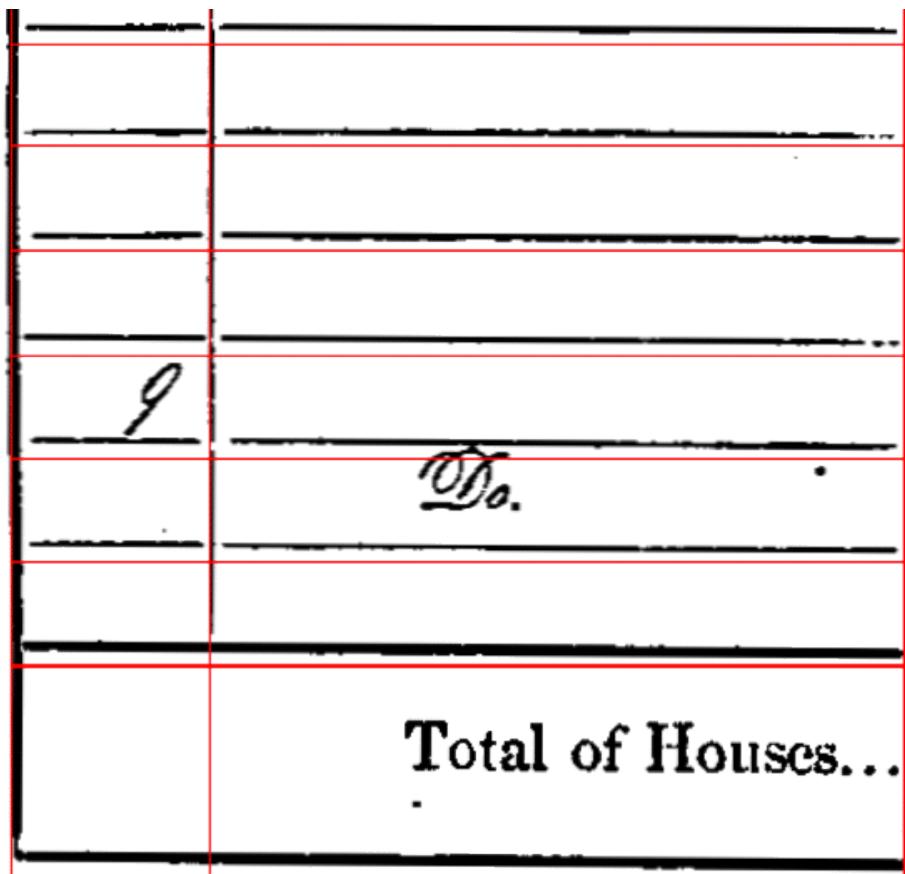
Initial Pass

The undermentioned Houses are situated within the Boundaries of the															
Civil Parish [or Township] of		City or Borough or District of		Municipal Ward of		Parliamentary Borough of		Town or Village or Hamlet of		Urban Sanitary District of		Rural Sanitary District of		Ecclesiastical Parish or District of	
St Edmund		Paisley		St. Mungo		Paisley		Paisley		Paisley		Paisley		Paisley	
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES In- habited (A), or Buildings (B)	NAME and Surname of each Person	RELATION to Head of Family	CON- DITION as to Marriage	AGE last Birthday of Male Females							Rank, Profession, or OCCUPATION	WHERE BORN	II (1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic
8 (reduced)	104b.		Charles Holloway Head Mrs. H. Dan Keppie —									Hartshorn			
9 30	Carlton Street 1		Deborah H d wife Mrs. Mary H. Macmillan Dairymaid									Carlton			
			Elijah d	Dairymaid	40	d	d	81				Paisley			
			Frances d	Dairymaid	38	d	d						d		
			William d	Dairymaid	17	S d	d						d		
			Frances d	Dairymaid	17	S d	d						d		
			William Gob	Postman	30	d	d						d		
10 44	Do	1	Elizabeth m Head Mrs. H. Dan Keppie —									Hartshorn			
			Elizabeth Lydia Dan Keppie										d		
			Jacob D. d	Dairymaid									d		
11 43	Do	1	Margaret Mary Do. d Dairymaid Dan Keppie —									Doonie Charlton			
			John Hart Head Mrs. H. Dan Keppie —									Stock Park			
			Elizabeth Hart Mrs. Mary Dan Keppie									Glencoe Park			
12 45	Do	1	Elizabeth Hart Dan Keppie —									Bellington			
			Charles Keppie Head Mrs. H. Dan Keppie —									Paisley			
			John Mary d wife Dan Keppie									Content			
			Hannah Chidzilimba d. M. 62 —									Douglas			
			John Mary Keppie Dan Keppie									Paisley			
			Isabel d d	d	d	10							d		
			Patricia d	Dairymaid	8								d		
			Georgina d	Dairymaid	7								d		
			Winfred d	Dairymaid	17								d		
			Laura d	Dairymaid	16								d		
Total of Houses	14	1	Total Males and Females	8	15										
Note. Draw the line through such of the words of the heading as are inappropriate.															

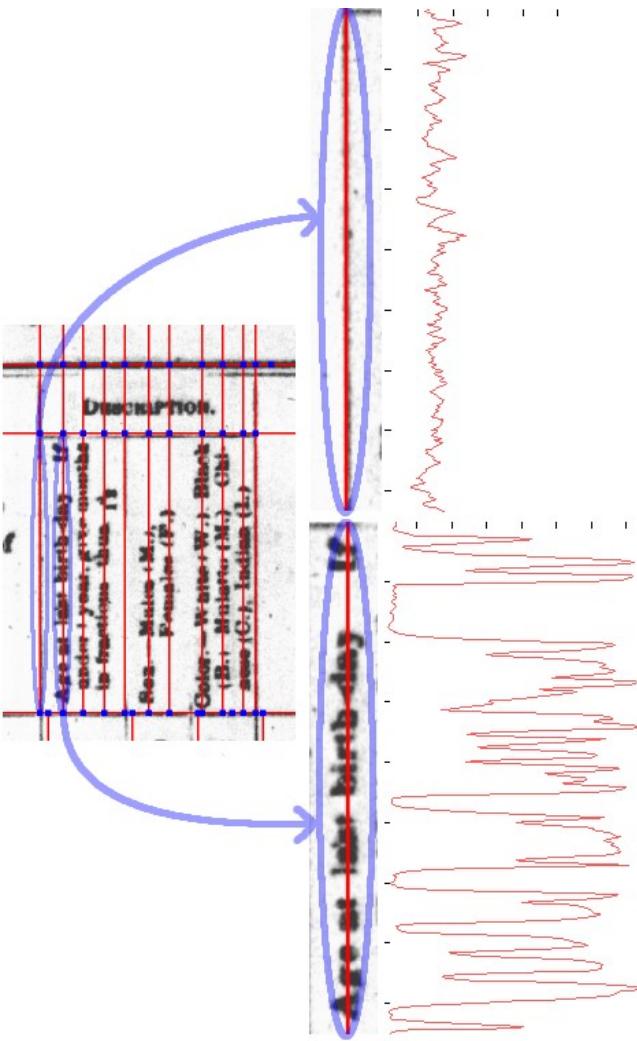
Image “Snapping”

- For each line segment in a row or column
 - Generate a profile over the segment’s area
 - Calculate line strength $ls_l(i) = \frac{1}{|i - gp| + 1} (ls_g) f(i)$
 - “Snap” to the location with the largest value

Image “Snapping”



False Positive Identification

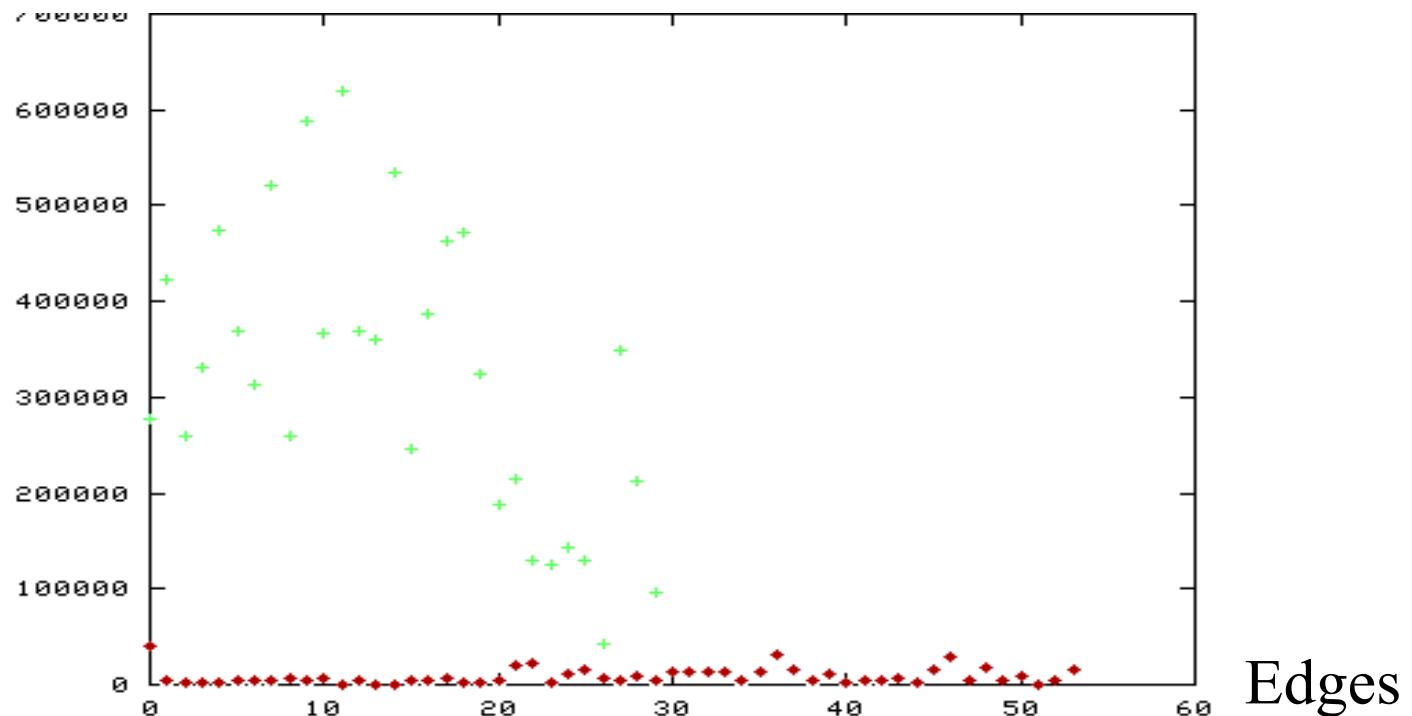


- Generate a profile perpendicular to the line segment
- Line profiles have low variance
- Text profiles have high variance

Edge Variance

The undermentioned Houses are situated within the Boundaries of the										
Civil Parish [or Township or District] of		City or Borough of	Municipal Ward of	Parliamentary Borough of	Town or Village or Hamlet of	Urban Sanitary District of	Rural Sanitary District of	Ecclesiastical Parish or District of		
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES	NAME and Surname of each Person	RELATION to Head of Family	CON- DITION as to Marriage	AGE last Birthday of Male [Female]	Rank, Profession, or OCCUPATION	WHERE BORN		
8 (Balanced)	10 Ch.	Unin- habited (U), or Building (B)								

Variance

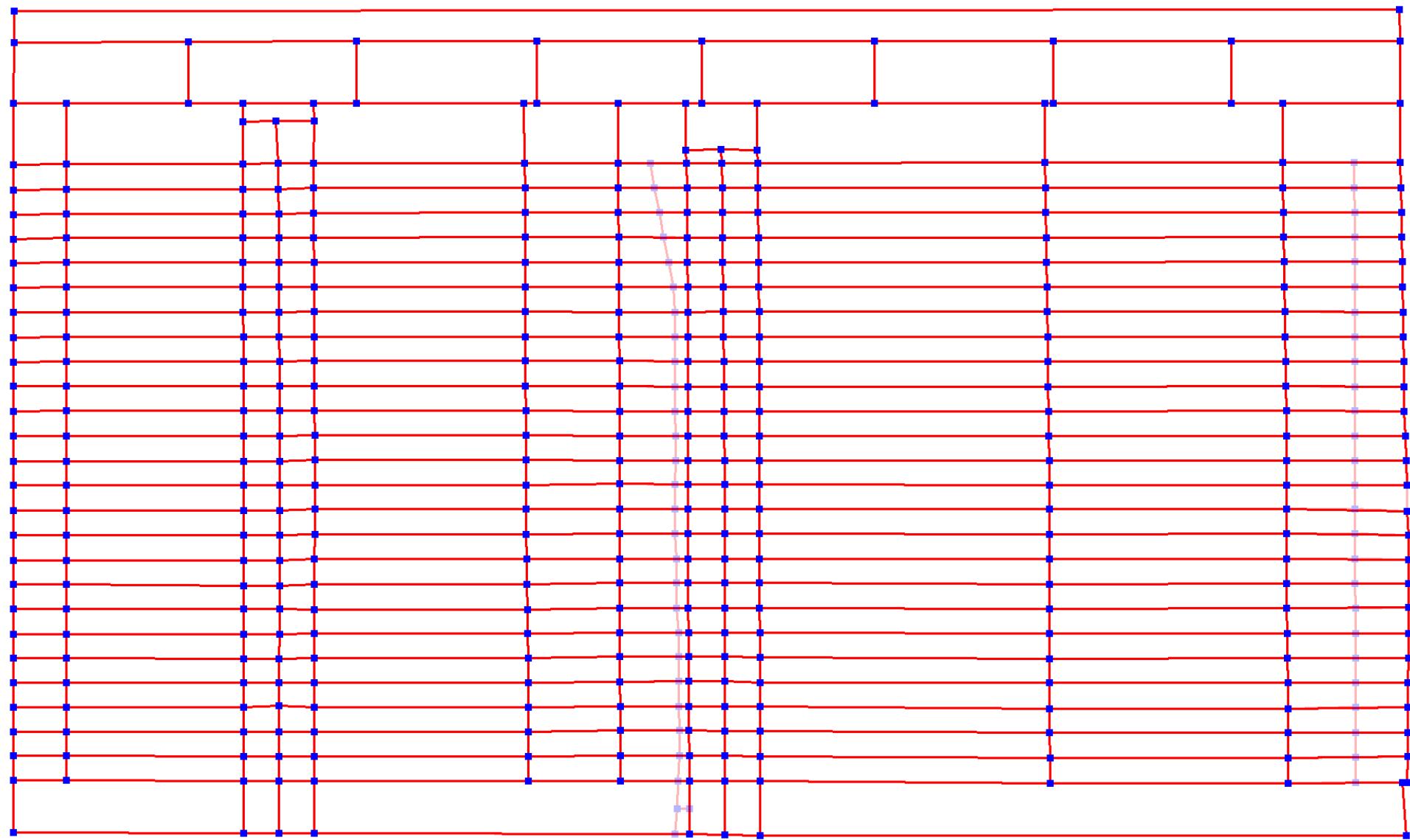


Document Template Creation

Inter-document Consensus

- Combine meshes generated from several documents
- Vote on line positions
- Discard line segments with a low vote count

Document Templates



Template to Image Registration

- Identify the document's body within the image
- Position the template to the corresponding location
- Locally snap each line segment to the image

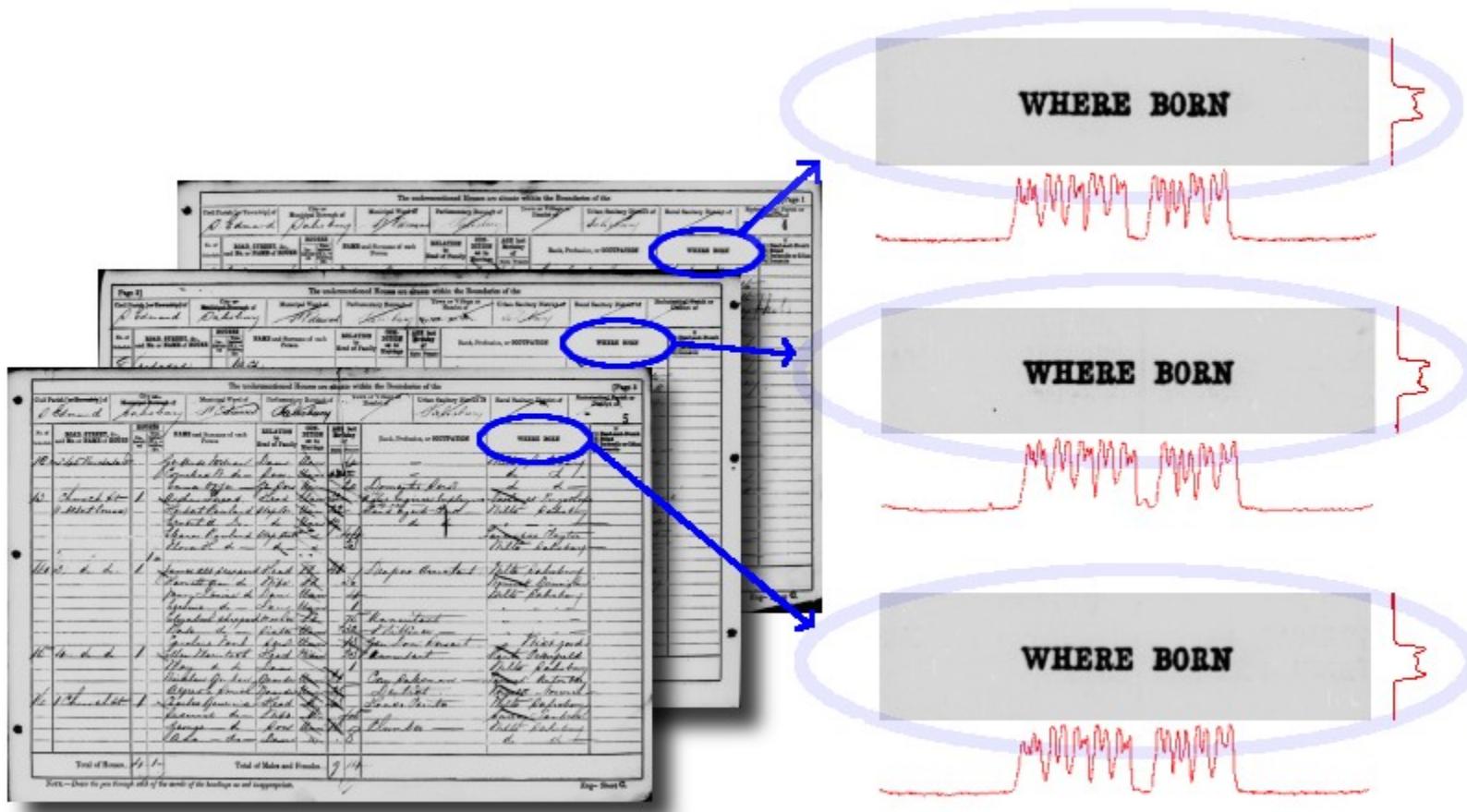
Template to Image Registration

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

Eng-Sheet D

Classification

Machine Printed Text



Classification

Handwriting



Zoned Image

Page 2]

The undermentioned Houses are situate within the Boundaries of the

Civil Parish [or Township] of P. Edmund.		City or Municipal Borough of Salisbury		Municipal Ward of P. Edmund		Parliamentary Borough of Salisbury		Town or Village or Hamlet of Salisbury		Urban Sanitary District of Salisbury		Rural Sanitary District of		Ecclesiastical Parish or District of		
No. of Schedule	ROAD, STREET, &c., and No. or NAME of HOUSE	HOUSES In- habit- ed	Un- in- habited (U.), or Building (B.)	NAME and Surname of each Person		RELATION to Head of Family	CON- DITION as to Marriage	AGE last Birthday of Males	AGE last Birthday of Females	Rank, Profession, or OCCUPATION		WHERE BORN		II (1) Deaf-and-Dumb (2) Blind (3) Imbecile or Idiot (4) Lunatic		
8	Salisbury	104 Ch.														
9	39 Nineharts	1		Charles Holloway	Head	X	X			Iron Worker		Southampton				
	" Castle Street			Deborah H. d.	Wife	X	X			Assistant Barber		Dorset Marshall				
				Eliza d.	Daughter	X	X	20		dr dr		Salisbury				
				Clementina d.	Daughter	X	X	18		dr dr		dr dr				
				William d.	Son	X	X	15		Painter		dr dr				
				Frances d.	Daughter	X	X	16		Painter		dr dr				
				William Gale	servt.	X	X	15		Painter		dr Collingwood				
	10 44 Do	1		Ashley John Wm	Head	X	X	30		Painter		Salisbury				
				Elizabeth Gale	Daughter	X	X	5		Painter		dr				
				Josie d.	Son	X	X	9		Painter		dr				
				Margaret Mayor	servt.	X	X	15		General Servt		Dorset Charlton Marshall				
11 45 Do	1			John Hart	Head	X	X	30		Cook		Stoke Park				
				Elizabeth Hart	Wife	X	X	23		Cook		Glossop North				
				Elizabeth Hart	Daughter	X	X	12		Cook		Shillingstone				
	12 45 Do	1		Charles Farmer	Head	X	X	35		Schoolmaster		Salisbury				
				John May, d.	Wife	X	X	32		Schoolmaster		Corton				
				Hannah Churchill	Daughter	X	X	62		Farmers Wife		Dorset Lympstone				
				Anna Mary Farmer	Daughter	X	X	12		Farmers Wife		Salisbury				
				Isabel d. d.	Daughter	X	X	10				dr				
				Walter d.	Son	X	X	8				dr				
				George d.	Daughter	X	X	7				dr				
				Walter d.	Daughter	X	X	7				dr				
				Laura d.	Daughter	X	X	6				dr				
Total of Houses...		41		Total of Males and Females...		8	15									

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

Future Work

- Implement mesh-to-mesh registration
- Classification through the use of document templates