



Thresholding of Text Documents

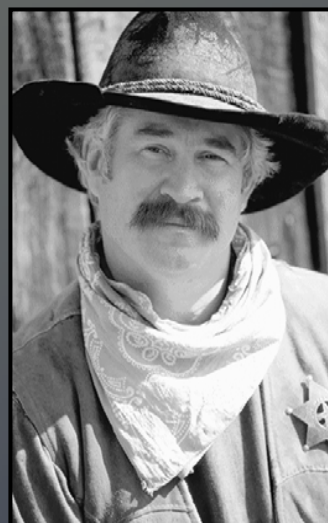
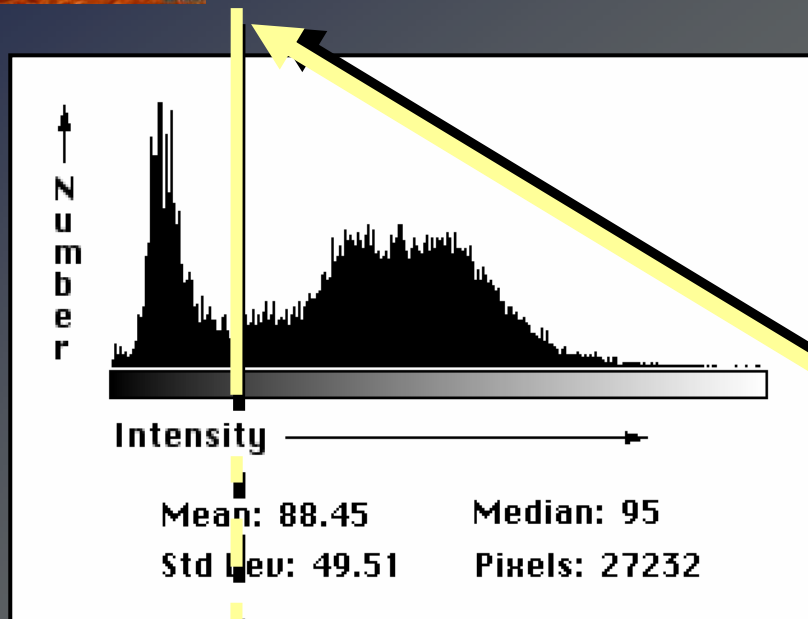
Oliver A Nina

William A Barrett



Thresholding or Binarization

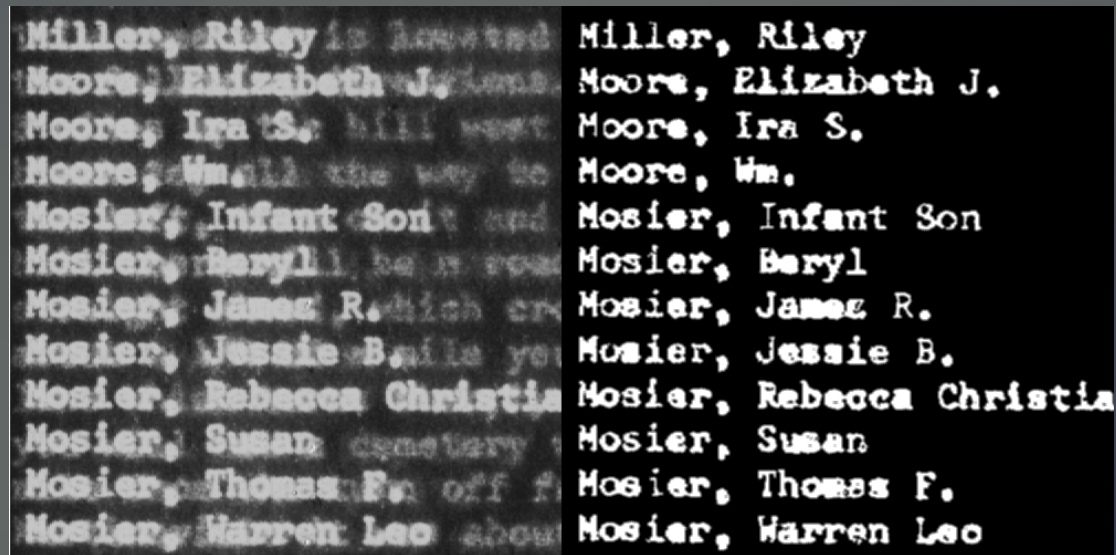
- Simple method of image segmentation
- The image is separated in two parts:
 - object of interest
 - background





Thresholding

– Important for the processing of scanned microfilms and OCR (Optical Character Recognition)



(Left) Original scanned record (Right) After Thresholding, Enhancement, and Antialiasing



The Problem

- Typical algorithms do a fairly well job on isolating the targeted object (text)
 - However, it is harder when the text looks similar to the background, such as with lighter pen strokes



- In many cases important pixels from the image are removed.



Previous Work

- ***Thresholding algorithm classification***
 1. Global Thresholding
 - 1.1 Bi-modal
 - 1.2 Multi-modal
 - 1.3 Multi-spectral
 2. Adaptive or Local Thresholding
 - 2.1 Hierarchical data structures
 - 2.2 Small window



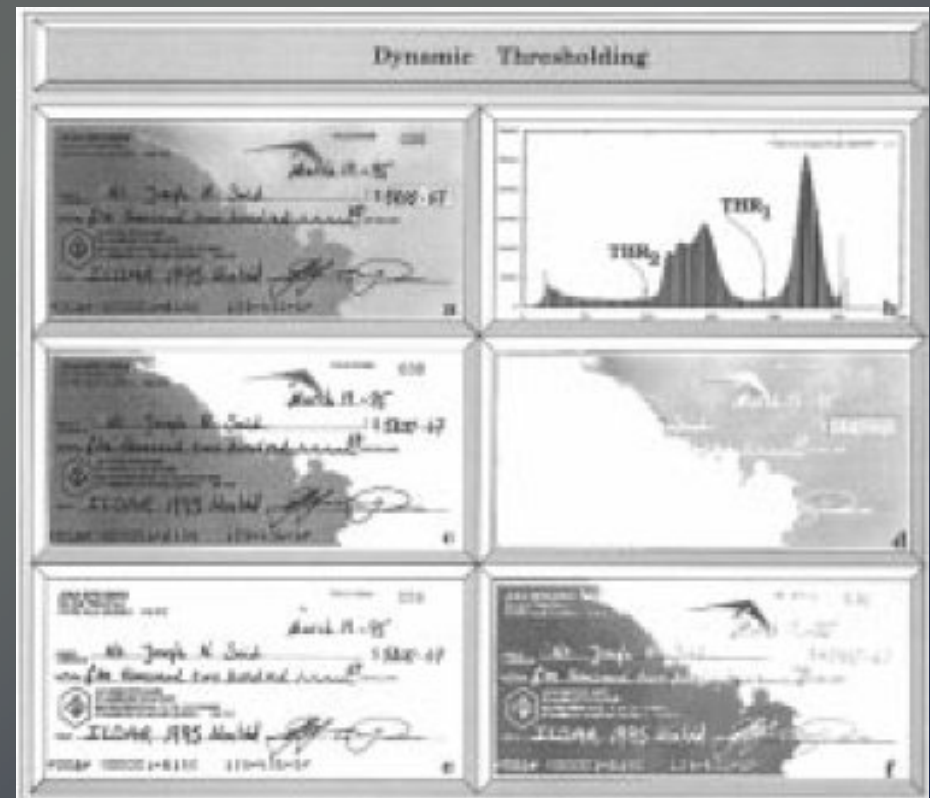
Thresholding Algorithms

- Examples of thresholding algorithms
 - Mean or Median value
 - Iterative Method
 - K-means
 - Otsu
 - Niblack
 - Yanowitz and Bruckstein



Related Work

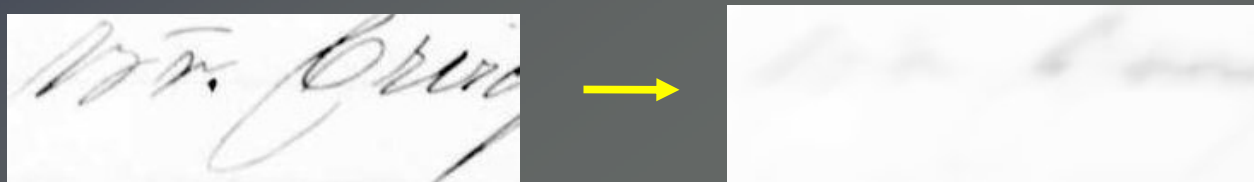
- Another similar recursive approach
 - *By Cheriet, Said, and Suen (June 1998)*
 - Used for bank checks
 - They use a training set to learn the background ($S=95\%$)
 - It only works if the targeted value is the darkest value in the image.



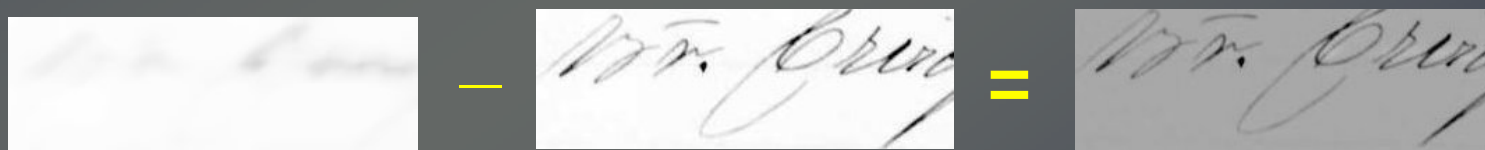


Our Approach “Rotsu”

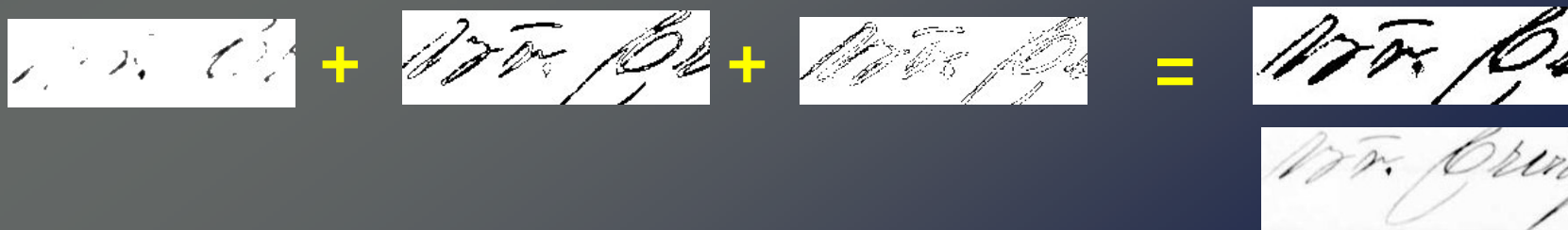
1. Background Estimation



2. Background Subtraction (Hutchinson 2004)



3. Apply Otsu Iteratively in different parts of the histogram

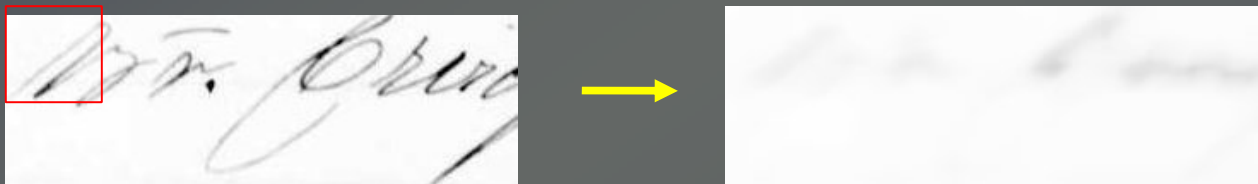




Our Approach

1. Estimation of Background

- We apply a median filter with a kernel of radius ~ 21 or bigger to the image



2. Background subtraction

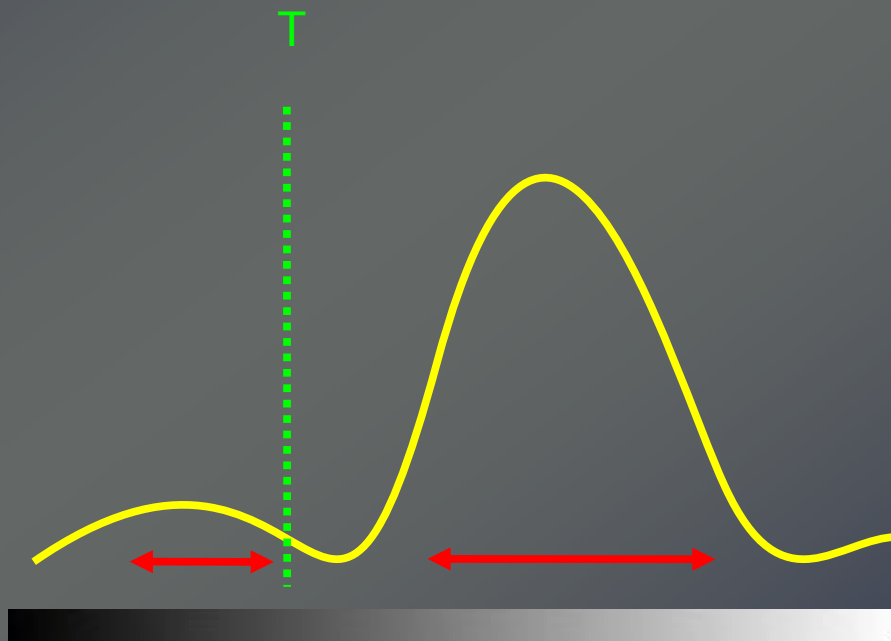
- We subtract the original image from the background
- We normalize the histogram in order to get rid of negative values and be able to see remaining pixels





Our Approach

3. The Otsu Algorithm

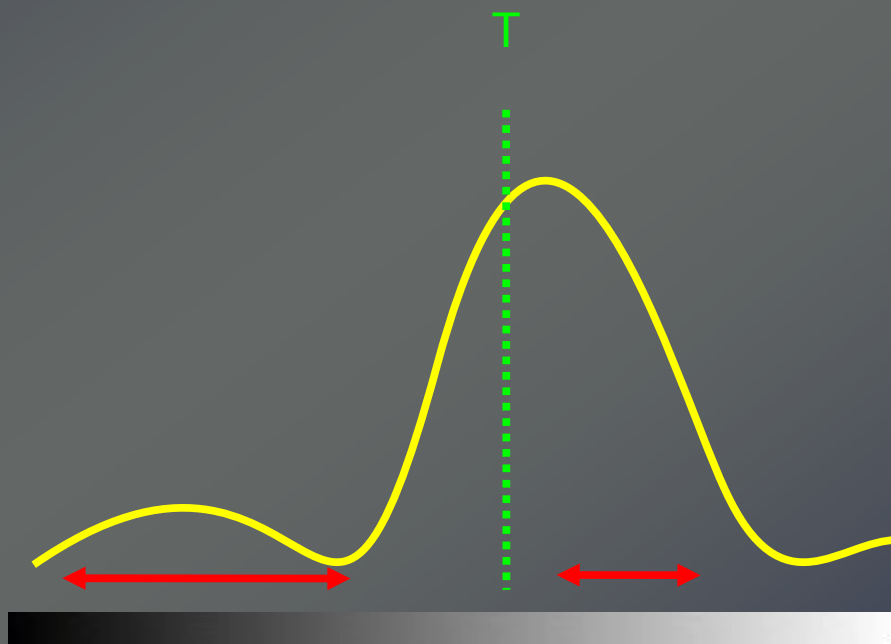


Goal: Minimize within variance class

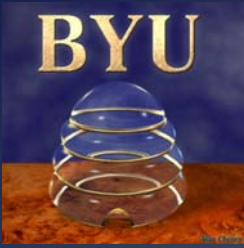


Our Approach

3. The Otsu Algorithm

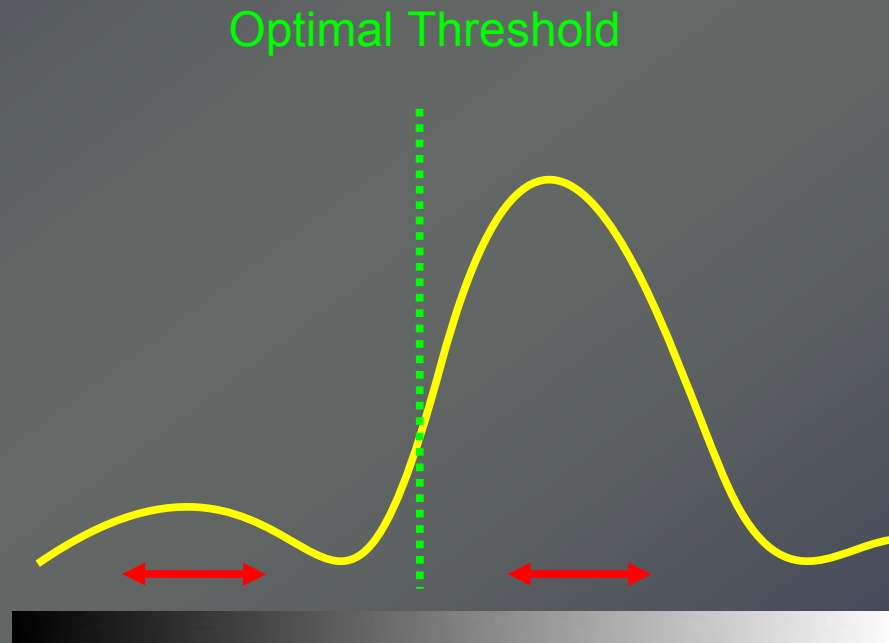


Goal: Minimize within variance class



Our Approach

3. The Otsu Algorithm



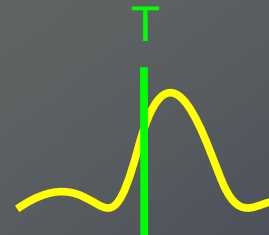
Goal: Minimize within variance class



Otsu

- Mathematically

$$\sigma^2 \text{ Within}(T) = nB(T)\sigma^2B(T) + nO(T)\sigma^2O(T)$$

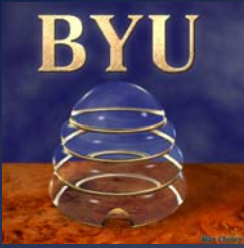


$$nB(T) = \sum_{i=0}^{T-1} p(i)$$

$\sigma^2B(T)$ = the variance of the pixels in the background (below threshold)

$$nO(T) = \sum_{i=T}^{N-1} p(i)$$

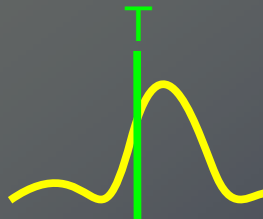
$\sigma^2O(T)$ = the variance of the pixels in the foreground (above threshold)



Otsu

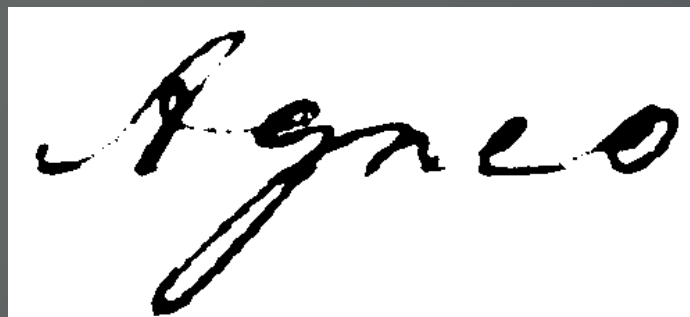
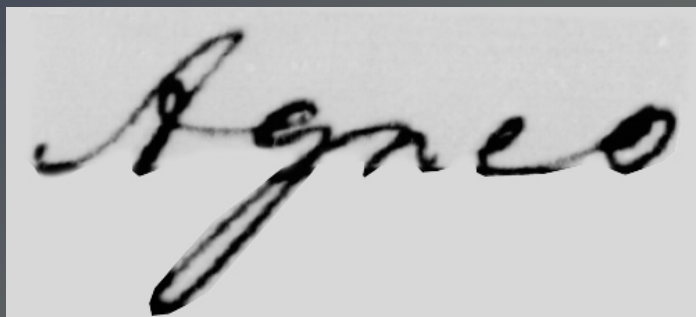
- Calculating within-class variance is too expensive
- Another way is to **maximize between-class variance**

$$\sigma^2 = \downarrow \sigma^2 \text{ Within}(T) + \sigma^2 \text{ Between}(T) \uparrow$$





Otsu





Rotsu

Recursive Otsu



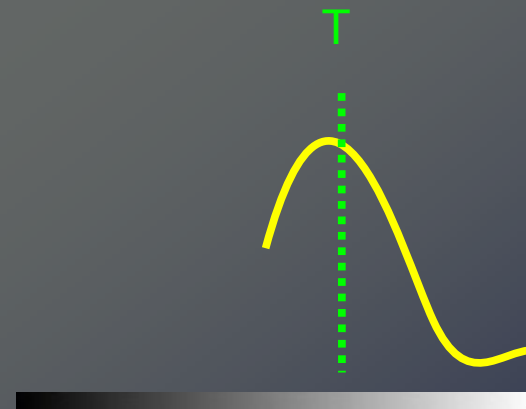
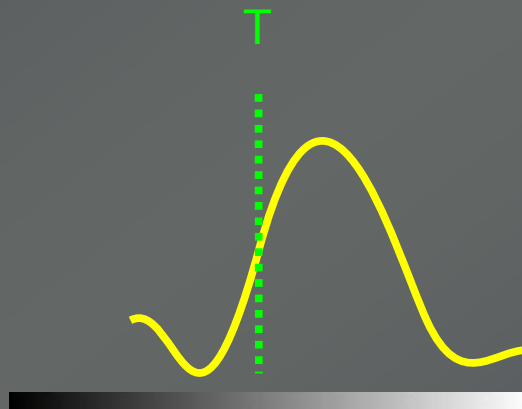
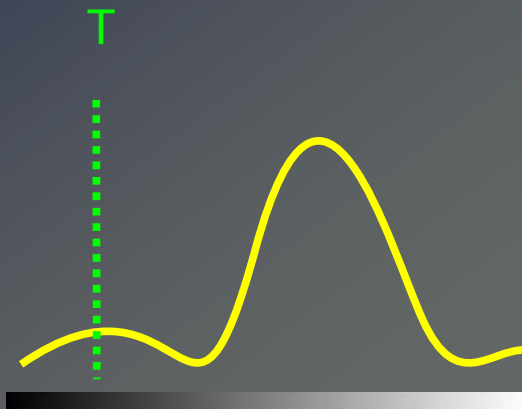
The algorithm

```
threshold = Otsu(image)
thresholdImage(image, thImg, threshold)

While(threshold < 255) { // until no more to threshold
    excludePixels(image, thImg, excludedImage)
    threshold = Otsu(excludedImage)
    thresholdImage(excludedImage, thImg, threshold)
    saveAndDisplayImage(newImg)
}
```



The algorithm

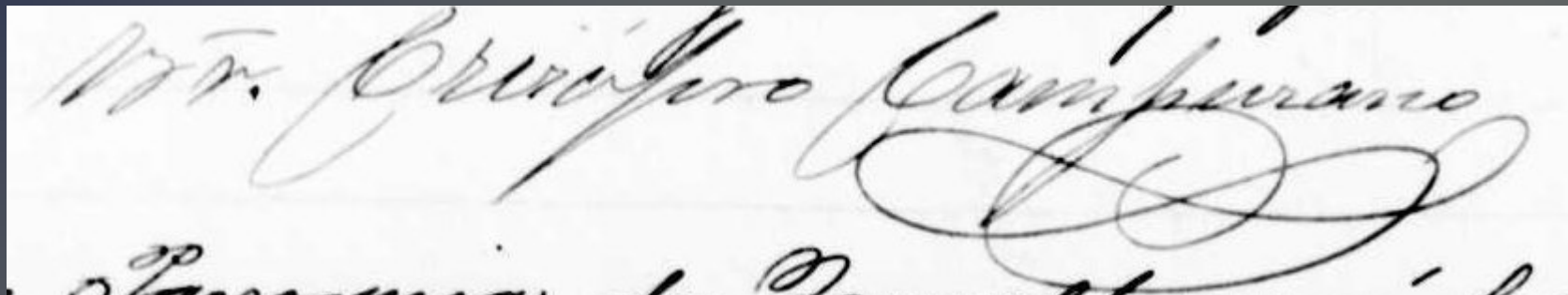




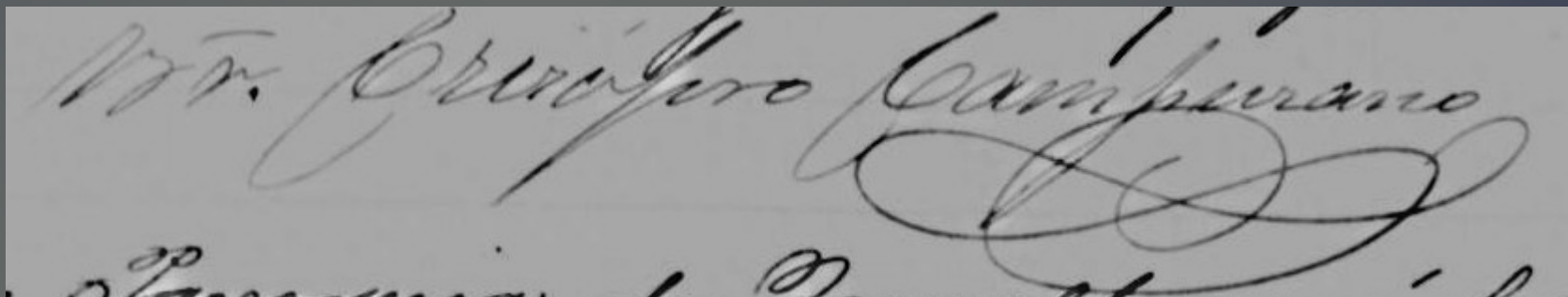
Results



Original Image

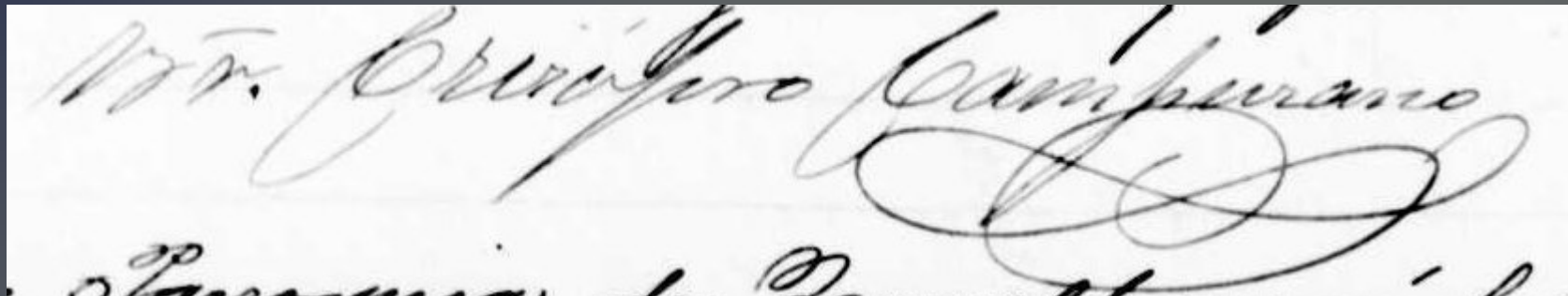


Original with background subtracted

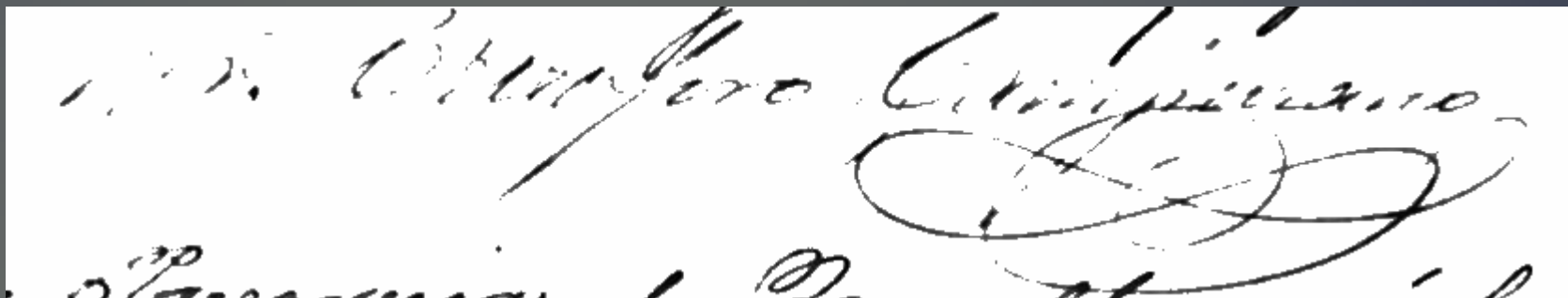




Original Image

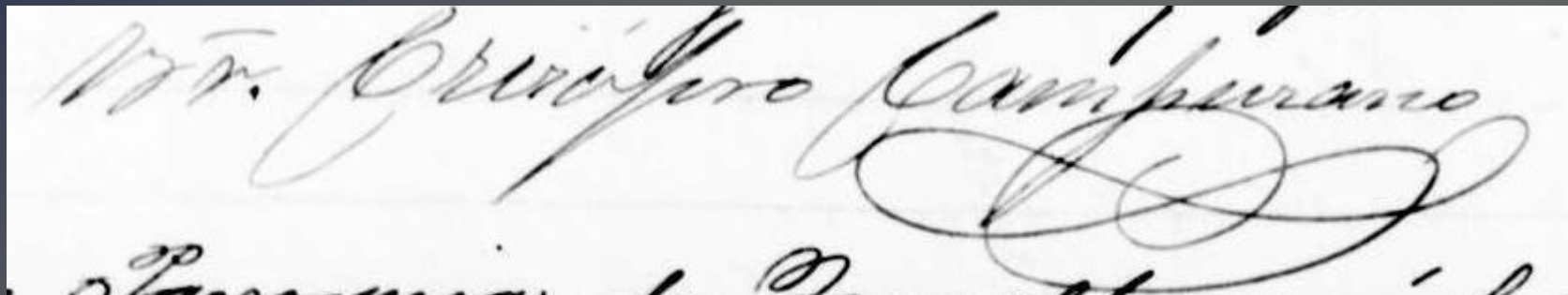


First Set = S1





Original Image

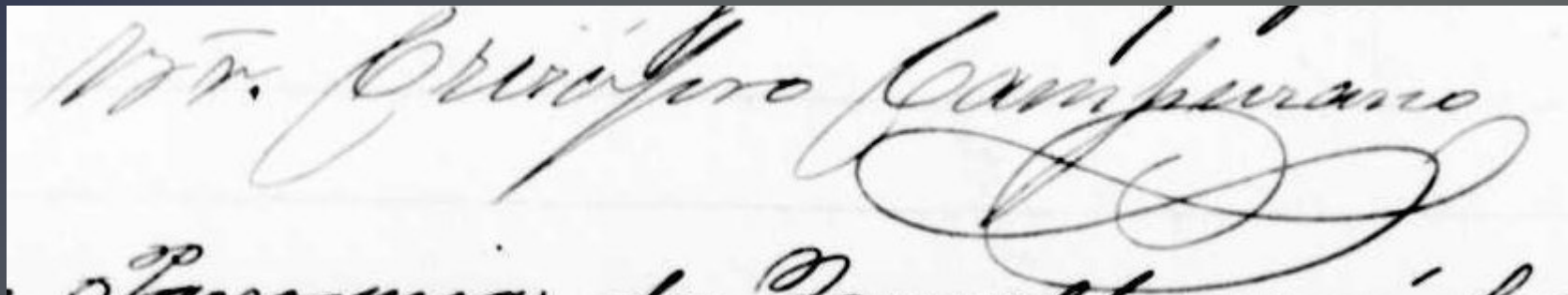


Second Set =S2

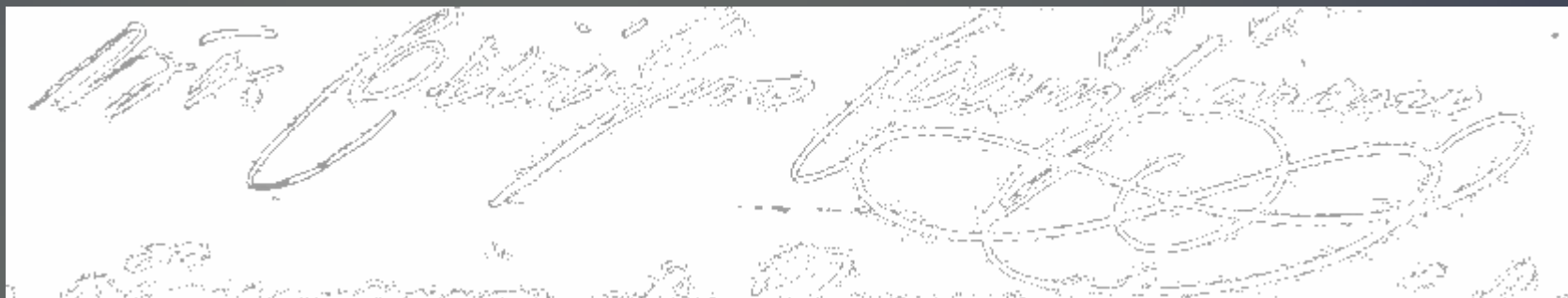




Original Image

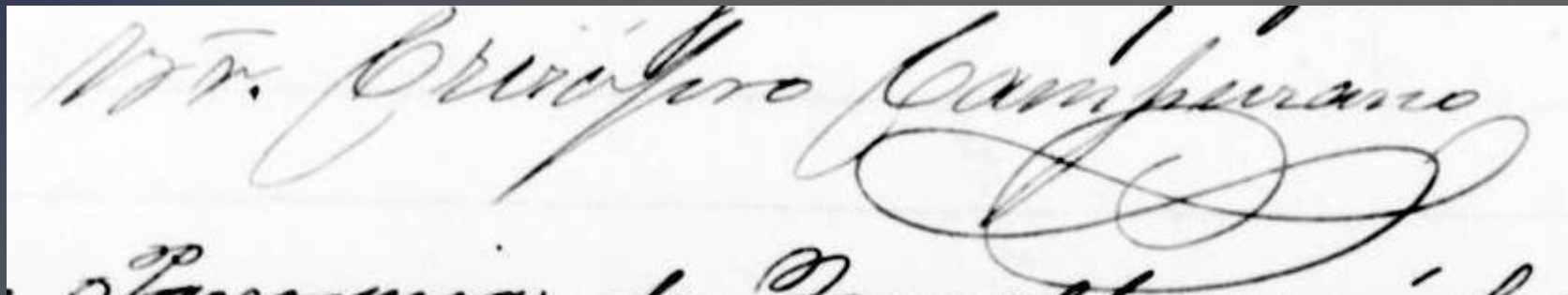


Third Set = S3

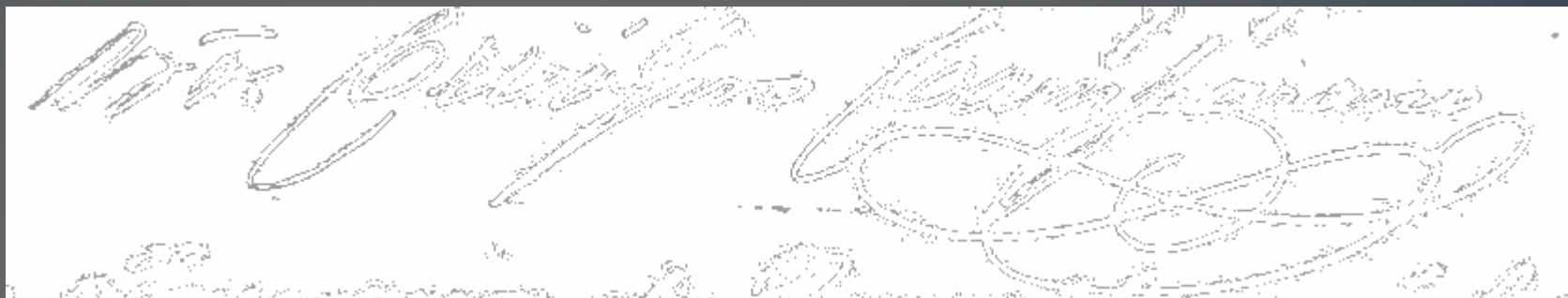




Original Image

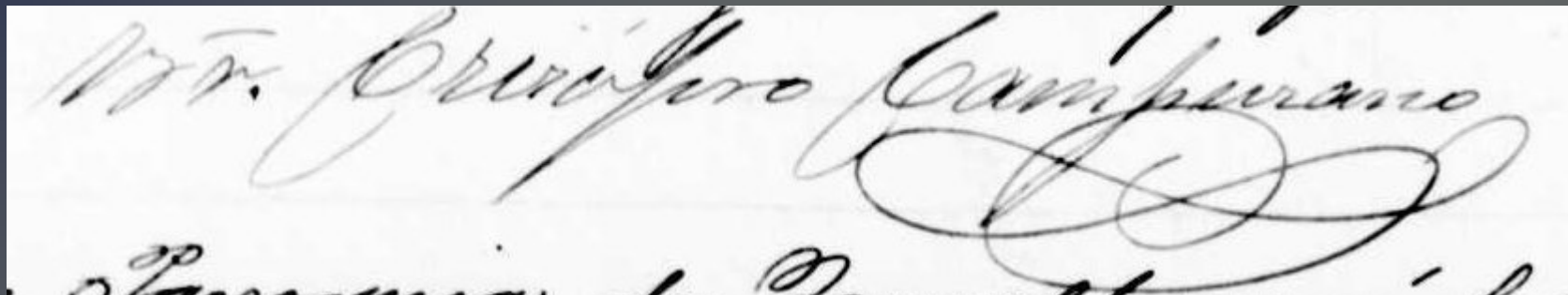


Fourth Set = S4

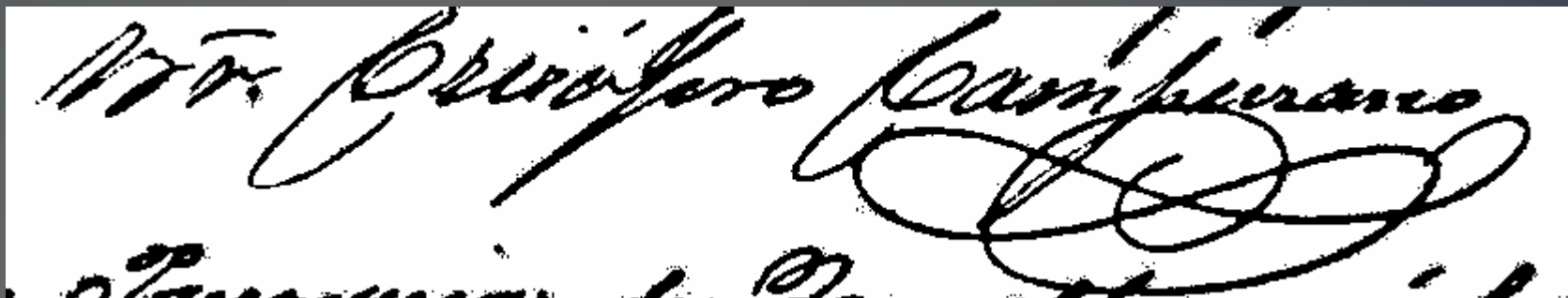




Original Image

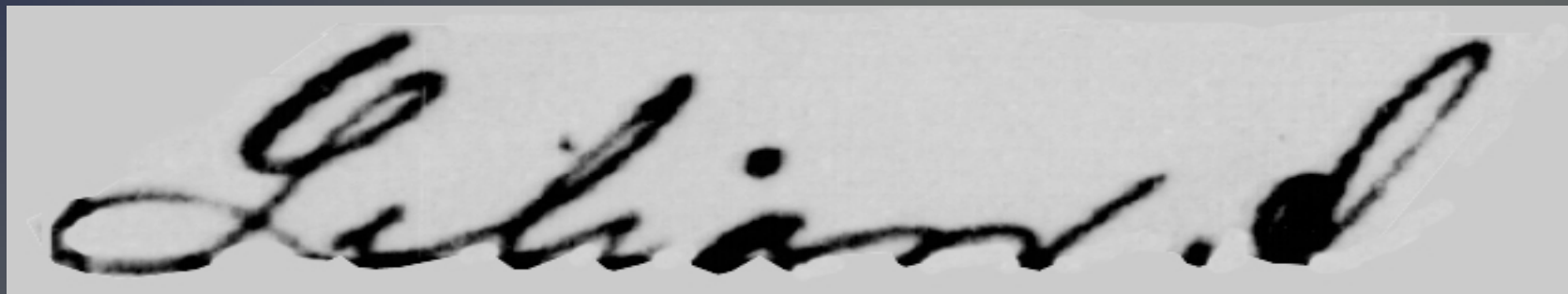


S1 + S2 + S3 + S4

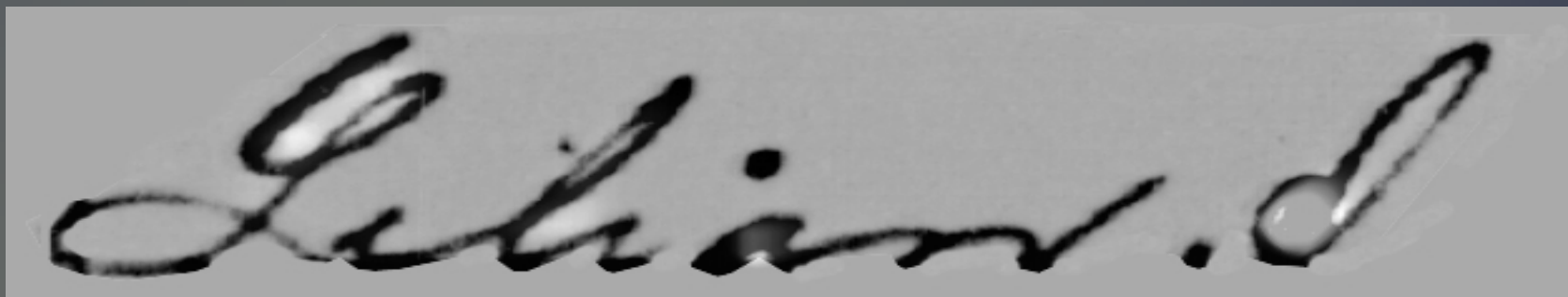




Original Image

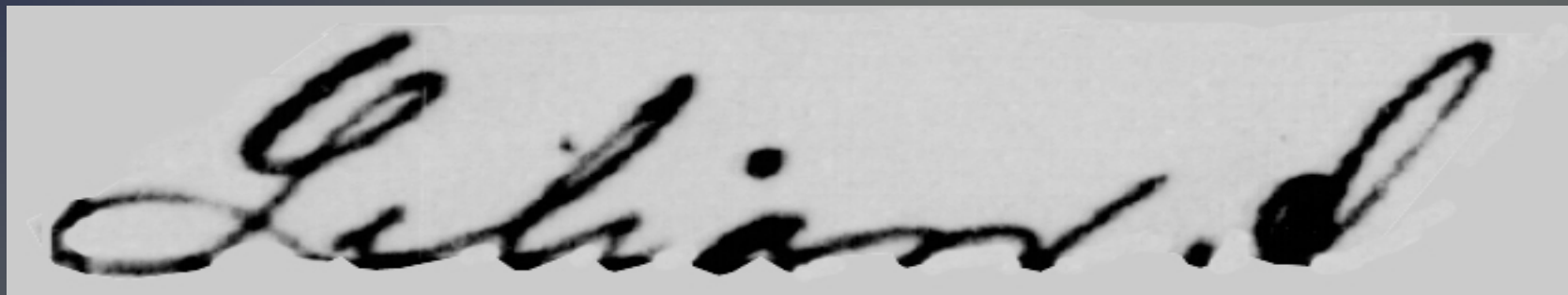


Original with background subtracted (K=41)

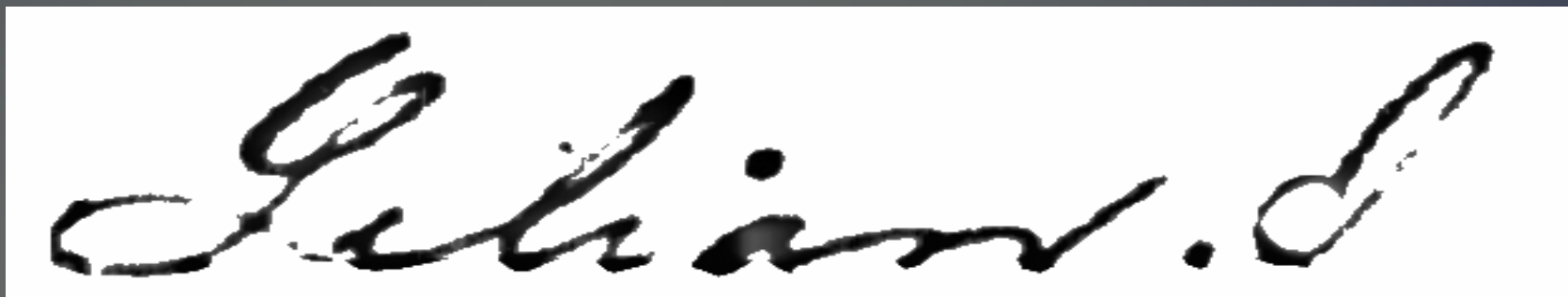




Original Image

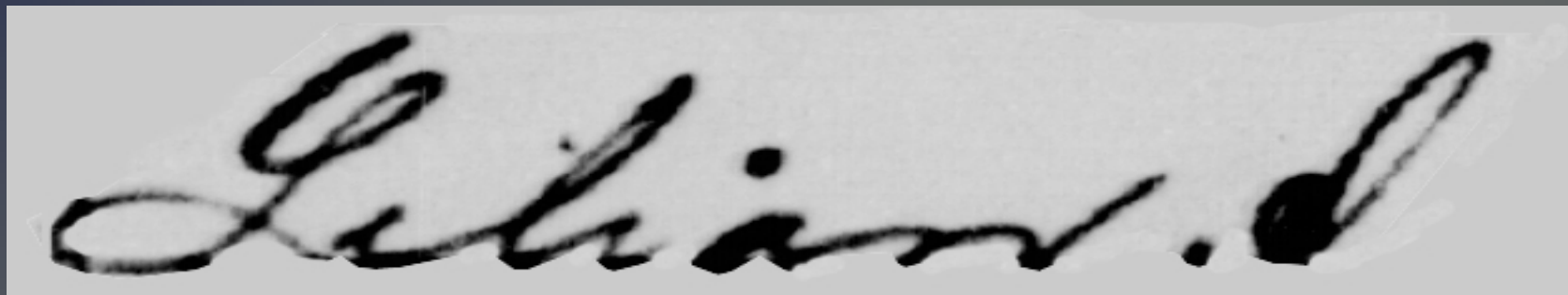


First Set =S1





Original Image

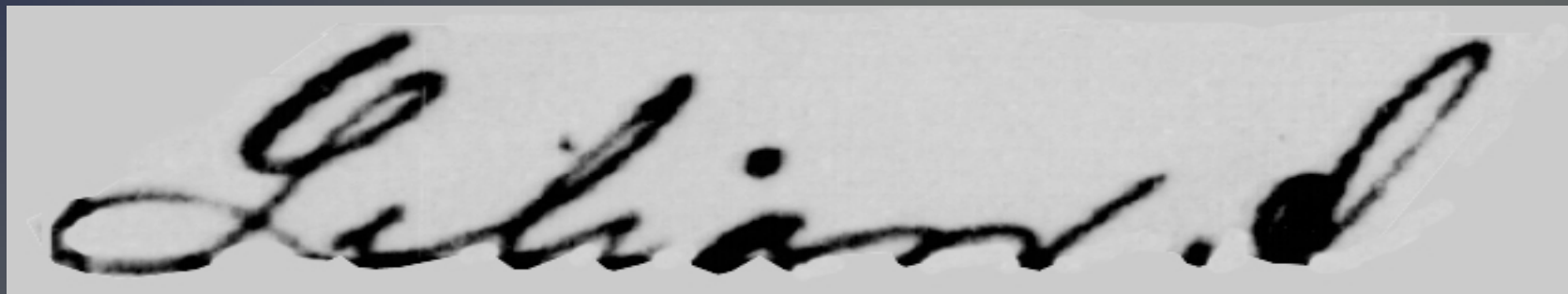


Second Set = S2

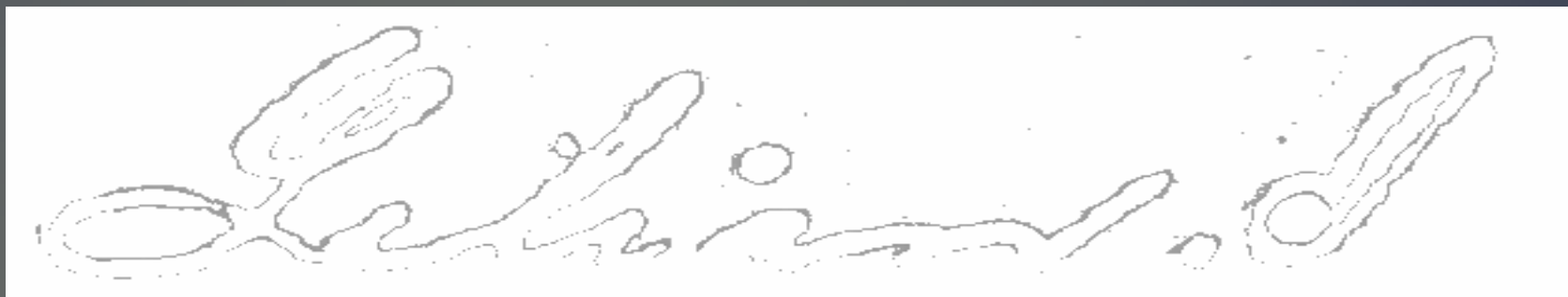




Original Image

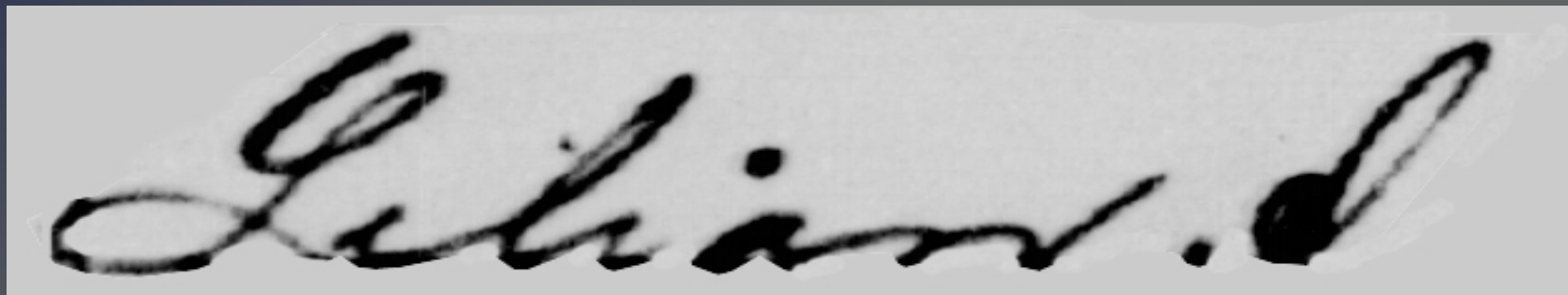


Third Set = S3





Original Image



$S_1 + S_2 + S_3$





Original Image

solamente a un
y le nombre Dionisio
Gurtin Dominguez y de
inos de este pueblo.
Salvador y y Teodora
pueblo a quienes ad

Background Approximation

solamente a un
y le nombre Dionisio
Gurtin Dominguez y de
inos de este pueblo.
Salvador y y Teodora
pueblo a quienes ad



Original Image

solamente a un
y le nombre Dionisio
Gustino Dominguez y et
inos de este pueblo.
salvador y y Teodora
pueblo a quienes ad

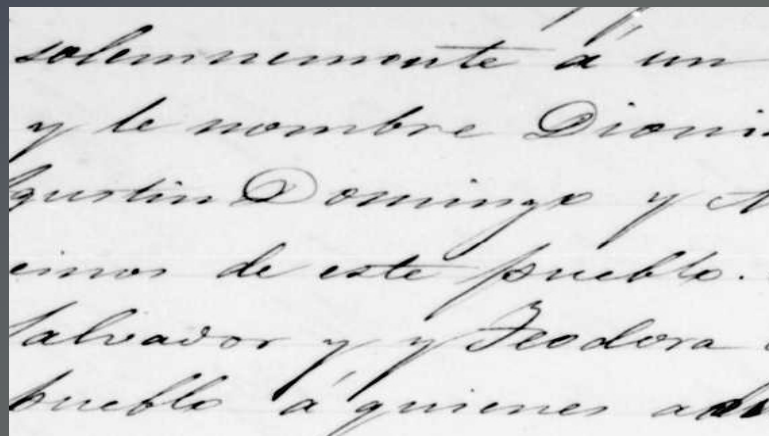
First Threshold = T_1

First Iteration

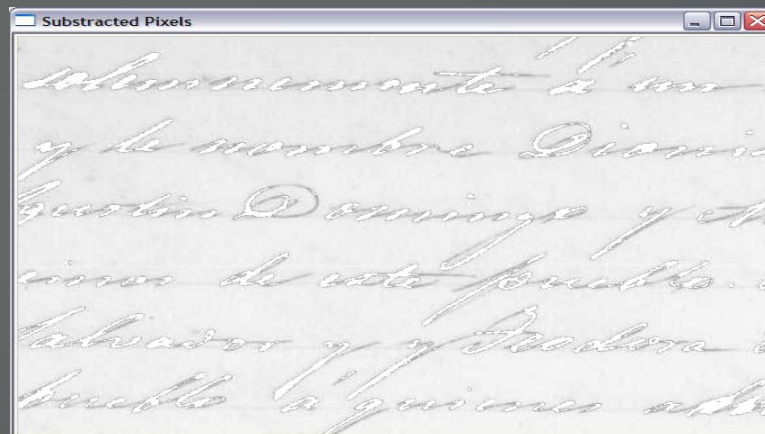
solamente a un
y le nombre Dionisio
Gustino Dominguez y et
inos de este pueblo.
salvador y y Teodora
pueblo a quienes ad



Original Image



Remaining Pixels

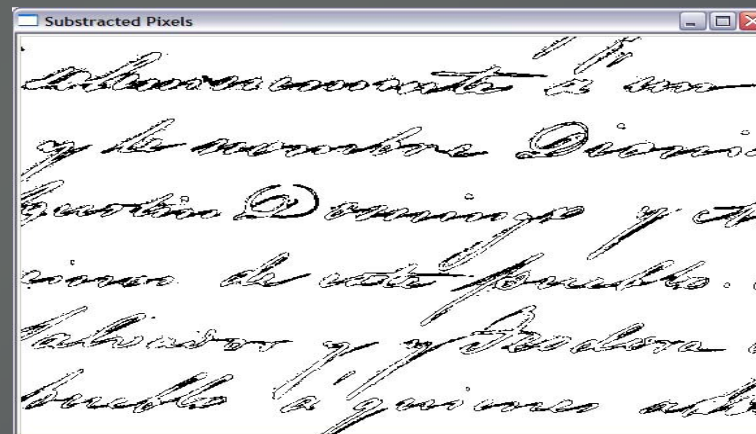




Original Image

solamente a un
y le nombre Dionisio
Gurtin Dominguez y de
inos de este pueblo.
salvador y y Feodora
pueblo a quienes ad

Second Threshold = T2

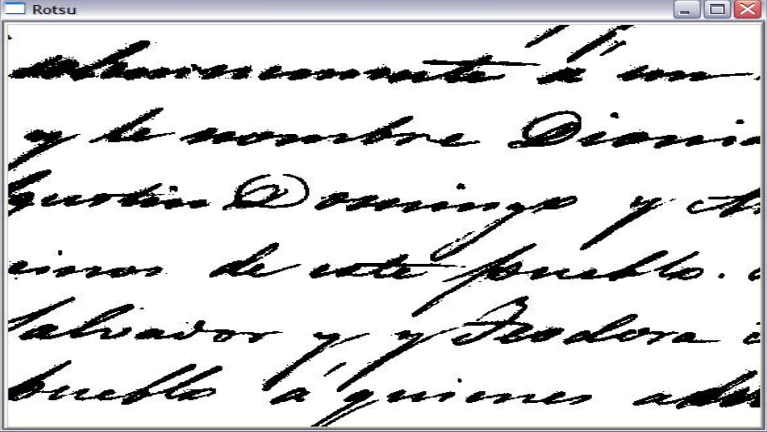




Original Image

solamente a un
y le nombre Dionisio
Gustin Dominguez y de
inos de este pueblo.
salvador y y Teodora
pueblo a quienes adu

T1 + T2



The image shows a computer window titled "Rotsu" containing a scan of the same handwritten text as the "Original Image". The text is written in a cursive script and reads: "solamente a un", "y le nombre Dionisio", "Gustin Dominguez y de", "inos de este pueblo.", "salvador y y Teodora", "pueblo a quienes adu". The window has standard Windows-style controls (minimize, maximize, close) in the top right corner.

solamente a un
y le nombre Dionisio
Gustin Dominguez y de
inos de este pueblo.
salvador y y Teodora
pueblo a quienes adu



Page 6

M.S. in the Parish of Hope under Dismore in the
City of Hereford in the year One thousand
eight hundred and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
late Dale	Lerninster	Feb 10 th	37	W. Logatt.
ian Stafford	Hope (The stone)	April 12 th	94	W. Logatt.
ward Gough.	Woodmanthorpe	April 16 th	76	W. Logatt.
James Smith.	Hill Hole	April 24 th	100	W. Andrews.
Thomas Grafton	Newton.	May 1 st	67	W. Andrews.
James Goodwin Quinn	The little Hampton Park	June 15 th	64	W. Andrews.
John Mellick's	The Pipers House Agar Hill	June 19 th	10	W. Andrews.
Elizabeth Bather	The Vaillets Agar Hill	June 29 th	60	W. Logatt.

Original Image

Page 6

M.S. in the Parish of Hope under Dismore in the
City of Hereford in the year One thousand
eight hundred and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
late Dale	Lerninster	Feb 10 th	37	W. Logatt.
ian Stafford	Hope (The stone)	April 12 th	94	W. Logatt.
ward Gough.	Woodmanthorpe	April 16 th	76	W. Logatt.
James Smith.	Hill Hole	April 24 th	100	W. Andrews.
Thomas Grafton	Newton.	May 1 st	67	W. Andrews.
James Goodwin Quinn	The little Hampton Park	June 15 th	64	W. Andrews.
John Mellick's	The Pipers House Agar Hill	June 19 th	10	W. Andrews.
Elizabeth Bather	The Vaillets Agar Hill	June 29 th	60	W. Logatt.

Background Subtracted



Page 6

U.S. in the Parish of Hope under Denmark in the
city of Herndon in the year One thousand
and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
late Dale	Lennister	Feb 10 th	37	W. Loggatt
John Stafford	Hope (The Home)	April 12 th	94	W. Loggatt
Edward Gough	Woodman's	April 16 th	76	W. Loggatt
James Smith	Hill Hole	April 20 th	100	W. Andrews
Thomas Grafton	Newton	May 1 st	67	W. Andrews
James Gordon Quinn	The Hill Hampton Park	June 15 th	64	W. Andrews
John Mellick	The Poplar Hill Agar Hill	June 19 th	10	W. Andrews
Elizabeth Baines	The Valletts Agar Hill	June 29 th	60	W. Loggatt

Original Image

Page 6

U.S. in the Parish of Hope under Denmark in the
city of Herndon in the year One thousand
and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
late Dale	Lennister	Feb 10 th	37	W. Loggatt
John Stafford	Hope (The Home)	April 12 th	94	W. Loggatt
Edward Gough	Woodman's	April 16 th	76	W. Loggatt
James Smith	Hill Hole	April 20 th	100	W. Andrews
Thomas Grafton	Newton	May 1 st	67	W. Andrews
James Gordon Quinn	The Hill Hampton Park	June 15 th	64	W. Andrews
John Mellick	The Poplar Hill Agar Hill	June 19 th	10	W. Andrews
Elizabeth Baines	The Valletts Agar Hill	June 29 th	60	W. Loggatt

S1

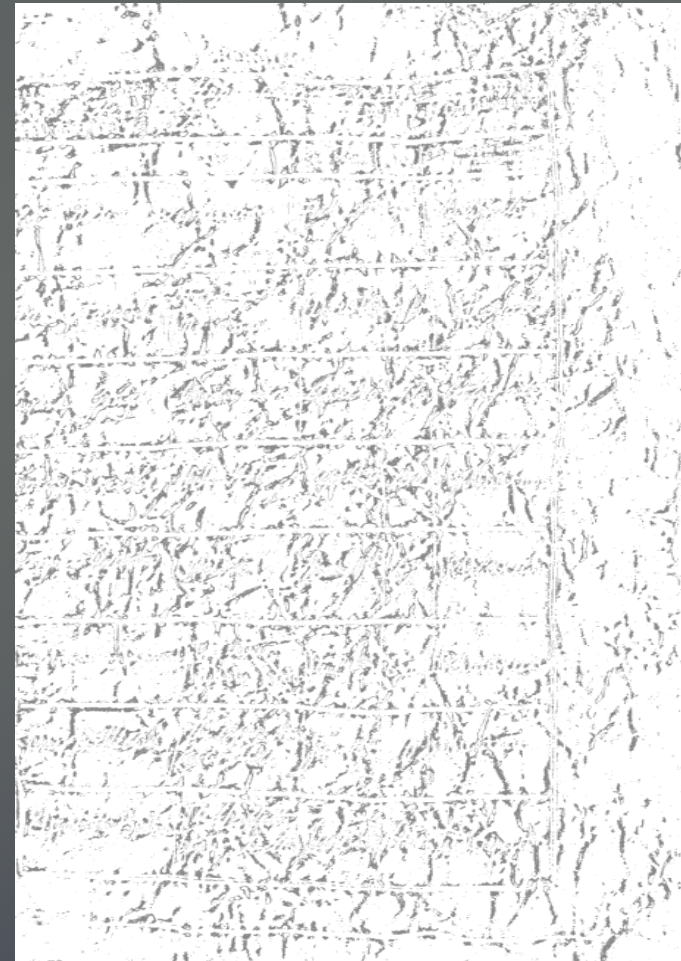


Page 6

U.S. in the Parish of Hope under Denmark in the
 city of Herford in the year One thousand
 eight hundred and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
John Dale	Lennister	Feb 10 th 37		W. Loggatt.
John Stafford	Hope (The Stone)	April 12 th 94		W. Loggatt.
Edward Gough.	Woodmanthorpe	April 10 th 76		W. Loggatt.
James Smith.	Hill Hole	April 20 th 100		W. Andrews.
Thomas Grafton	Newton.	May 1 st 67		W. Andrews.
James Goodwin Quinn	The Village Hampton Park	June 15 th 64		W. Andrews.
John Mellicham	The Poplar House Rogers Hill	June 19 th 10		W. Andrews.
Elizabeth Batham	The Vallyetto Rogers Hill	June 29 th 60		W. Loggatt.

Original Image



S3



Page 6

U.S. in the Parish of Hope Under Denmark in the
 City of Herford in the year One thousand
 eight hundred and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
John Dale	Lennister	Feb 10 th	37	W. Loggatt.
John Stafford	Hope (The Stone)	April 12 th	94	W. Loggatt.
Edward Gough.	Woodmanthorpe	April 16 th	76	W. Loggatt.
James Smith.	Hill Hole	April 24 th	100	W. Andrews.
Thomas Grafton	Newton.	May 1 st	67	W. Andrews.
James Goodwin Quinn	The Village Hampton Park	June 15 th	64	W. Andrews.
John Mellichamp	The Pipers House Rogers Hill	June 19 th	10	W. Andrews.
Elizabeth Batham	The Vallyetts Upper Hill	June 29 th	60	W. Loggatt.

Original Image



S3



Page 6

M.S. in the Parish of Hope Under Denmore in the
City of Hereford in the year One thousand
and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
John Dale	Lennister	Feb 12 th	37	W. Logatt
John Stafford	Hope (The Stone)	April 12 th	94	W. Logatt
Edward Gough	Woodmanthorpe	April 16 th	76	W. Logatt
James Smith	Hill Hole	April 24 th	100	W. Andrews
Thomas Grafton	Newton	May 1 st	67	W. Andrews
James Goodwin Quinn	The Hill Hampton Park	June 15 th	64	W. Andrews
John Mellick	The Pigeon House Agar Hill	June 19 th	10	W. Andrews
Elizabeth Bather	The Vallyetto Agar Hill	June 29 th	60	W. Logatt

Original Image

Page 6

M.S. in the Parish of Hope Under Denmore in the
City of Hereford in the year One thousand
and Eighty.

Name	Abode.	When buried.	Age.	By whom the Ceremony was performed.
John Dale	Lennister	Feb 12 th	37	W. Logatt
John Stafford	Hope (The Stone)	April 12 th	94	W. Logatt
Edward Gough	Woodmanthorpe	April 16 th	76	W. Logatt
James Smith	Hill Hole	April 24 th	100	W. Andrews
Thomas Grafton	Newton	May 1 st	67	W. Andrews
James Goodwin Quinn	The Hill Hampton Park	June 15 th	64	W. Andrews
John Mellick	The Pigeon House Agar Hill	June 19 th	10	W. Andrews
Elizabeth Bather	The Vallyetto Agar Hill	June 29 th	60	W. Logatt

S1 + S2 + S3



Henry, George G. Bjwate, David M. Davies,
 Thomas W. Rice, John Griffiths, John Gibbs, Rich-
 ard Palmer, Edward T. Edwards, and
 Joseph W. Morgan there. They had come there ac-
 cording to my directions. The General Council
 of all the Ministry in the British Mission commenced
 at 10. a. m., and these brethren had come to Birning-
 ham for this purpose. Knowing bro. & sister Smith, as
 being also aware of the exorbitant charges generally
 made for lodgings in all strange places, and the low
 circumstances of my brethren, I wrote to bro. & sister
 Smith some three weeks before hand, requesting them to
 endeavour to obtain lodgings for the brethren at as low a
 charge as possible. They managed to find them lodgings
 at a decent public-house at 4d. each per night, and
 I managed to get bro. & sister Smith's consent to let them
 eat their own food at their house in the morning
 and afternoon, by which the brethren were enabled to
 have sufficient of bread and butter and tea and
 coffee at less than 3d. per meal, besides the
 comfort of having their food at a place where they
 could feel free and unfettered. When I was at
 the Council that was held here at the commence-
 ment of 1859, I had to pay 1/- each ^{night} for my lodgings,
 and very heavy price for food. This was not very ex-
 orbitant, yet it was three times the sum the brethren
 paid this time, and yet not so comfortable.
 After I had a little repast, I repaired to Farm-stre-
 et Chapel, Hockley, where the Council was being held,
 the minutes of which I shall hereafter paste. This
 afternoon's meeting commenced at 4 and terminated at

Original Image

Henry, George G. Bjwate, David M. Davies,
 Thomas W. Rice, John Griffiths, John Gibbs, Rich-
 ard Palmer, Edward T. Edwards, and
 Joseph W. Morgan there. They had come there ac-
 cording to my directions. The General Council
 of all the Ministry in the British Mission commenced
 at 10. a. m., and these brethren had come to Birning-
 ham for this purpose. Knowing bro. & sister Smith, as
 being also aware of the exorbitant charges generally
 made for lodgings in all strange places, and the low
 circumstances of my brethren, I wrote to bro. & sister
 Smith some three weeks before hand, requesting them to
 endeavour to obtain lodgings for the brethren at as low a
 charge as possible. They managed to find them lodgings
 at a decent public-house at 4d. each per night, and
 I managed to get bro. & sister Smith's consent to let them
 eat their own food at their house in the morning
 and afternoon, by which the brethren were enabled to
 have sufficient of bread and butter and tea and
 coffee at less than 3d. per meal, besides the
 comfort of having their food at a place where they
 could feel free and unfettered. When I was at
 the Council that was held here at the commence-
 ment of 1859, I had to pay 1/- each ^{night} for my lodgings,
 and very heavy price for food. This was not very ex-
 orbitant, yet it was three times the sum the brethren
 paid this time, and yet not so comfortable.
 After I had a little repast, I repaired to Farm-stre-
 et Chapel, Hockley, where the Council was being held,
 the minutes of which I shall hereafter paste. This
 afternoon's meeting commenced at 4 and terminated at

S1

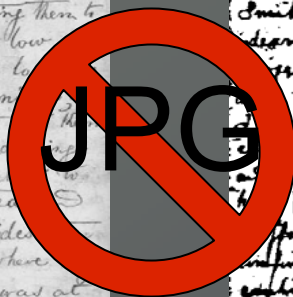


Jeremy, George G.,
Thomas W. Rees, J.
and Palmer, Ed.
Joseph W. Morgan

James, George G.
Thomas M. Reed, J.
and Palmer, Ed.
Frank W. Moore

James, George G. Biquette, David M. Davis, Thomas M. Rice, John Griffiths, John Gibbs, Richard Palmer, Edward T. Edwards, and Joseph M. Morgan there. They had come there according to my directions. The General Council of all the Mission in the British Mission arrived at 10. a. m. and these brethren had come to Birmingham for this purpose. Knowing, bro. & sister Smith, as being also aware of the exorbitant charges generally made for lodgings in all strange places, and the low circumstances of my brethren, I wrote to bro. & sister Smith some three weeks before hand, requesting them to endeavour to obtain lodgings for the brethren at as low charge as possible. They managed to find them by at a decent public-house at 4d. each per night. I managed to get bro. & sister Smith's consent to eat their own food at their house in the morning and afternoon, by which the brethren were enabled to have sufficient of bread and butter and tea & coffee at less than 3d. per meal, besides the comfort of having their food at a place where could feel free and unfettered. When I was at the Council that was held here at the commencement of 1859, I had to pay 1/- each ^{right} for my lodgings, and very heavy price for food. This was not an exorbitant, yet it was three times the sum the brethren paid this time, and yet not so comfortable. After I had a little repast, I repaired to Farm-street Chapel, Hockley, where the Council was being held, the minutes of which I shall hereafter state. This afternoon's meeting commenced at 4 and terminated at

among George G. Biesta, Daniel M. Jones,
Thomas W. Reed, John Griffiths, John Little, Rich-
ard Palmer, Edward J. Edwards, and
Joseph W. Morgan there. They had come there ac-
cording to my directions. The General Council
of all the Ministers in the British Mission Com-
mittee, as well as their brethren, had come to Bir-
mingham for this purpose. Knowing Mr. Deak Smith's
kindness and willingness to accommodate strangers generally
and his willingness to lodge in all strange places, and the low
circumstances of my brethren, I wrote to Mr. Deak Smith
asking that he would be so kind as to request them to
deavour to obtain lodgings for the brethren at as low a
rate as possible. The Committee for their lodgings
did not find it possible to find them lodgings
in a public house at 4d per night, and
I got to get a private house, and to let them
eat and food at their house in the morning
and afternoon only which the brethren were enabled to
sufficiently of bread and butter, and tea and
coffee and beer, and also for socials, besides their
expenses of having their food at a place where they
could feel fresh and comfortable. Then I was in-
formed by the Committee that we had been at the conven-
tion of 1859, I had to pay 1/6 each for my lodgings
and my daily food for food. This was not a case
of want, but of want, that time the same the brethren
paid their money and yet not so comfortable.
I felt that I had a little regret, I regretted the fact that
I had a little regret, I regretted the fact that
the Committee of which I shall be a member. This
afternoon I was very much at 4d and 1/6 per night at



Original Image

Final Composite



Conclusion

- Although Rotsu is still a work in progress, it definitely shows promising results
 - Rotsu allows us to save softer strokes that would be lost with conventional methods otherwise.
 - Relatively easy to implement.
 - Opens up the door to new ideas on how to improve thresholding.



Further Work

- Determine a better background estimate.
 - Automate the selection of kernel size for the median filter
 - Improve the criteria with which we decide to get rid of background pixels
 - Investigate to see if the combination of Rotsu with other techniques would be better



Questions?