

Writer Identification Using a Statistical And Model Based Approach



Diamantatos Paraskevas
Gritzalis Stefanos
Kavallieratou Ergina
University Of The Aegean

Presentation Overview

- ✓ Writer Identification-Objective
- ✓ Related work
- ✓ Skeleton Hinge distribution
- ✓ Proposed System
- ✓ Experimental Data & Results
- ✓ Conclusion

Writer Identification

Matching unknown handwritings against a database of samples with known authorship.

Objective



- ✓ Improvement of existing techniques
 - ✓ Edge-hinge distribution
 - ✓ Edge-hinge combinations
- ✓ Hypothesis: Stroke width has size 1.

Lucky indeed are the families that possess a heritage of lace. A lace bridal veil handed down from mother to daughters, a lace-trimmed Christening robe that each generation wears in turn - these are proud possessions, linked to the exquisite lace-making of the past.

We read, with sympathy, an advertisement in the reign of Charles 2 stating: 'Last: a lawn handkercher with a broad hem laced round with Fine Point lace about four fingers broad' and among the effects left by Nell Gwynn is an unpaid bill for 'scarlet satin shoes with silver lace'. The records of these tender trifles are very touching but then lace does place upon everything a delicate sentiment.

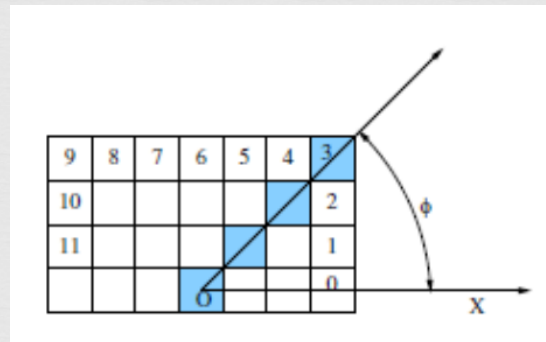
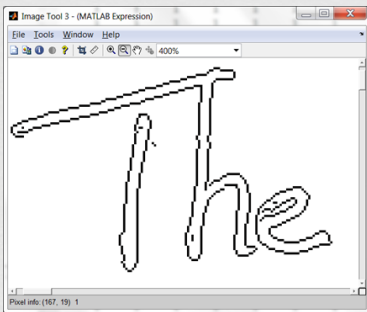
«Make everything as simple as possible, but not simpler.»

Albert Einstein

Edge-Direction Distribution*



- ✓ Edge detection.
- ✓ Direction of edge.
- ✓ Histogram of Directions.
- ✓ Normalized to a probability distribution.
- ✓ Nearest Neighbor.

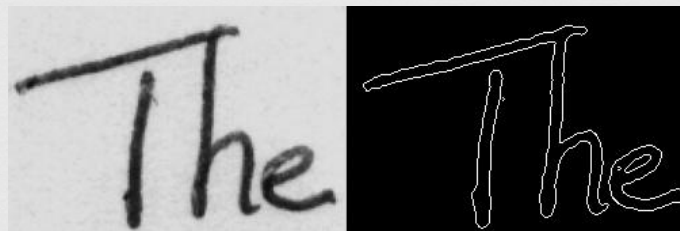


Accuracy: 35% for 250 writers.

Edge-Hinge Distribution*



- ✓ Statistical Feature.
- ✓ Outperforms all the other statistical approaches.
- ✓ Based on Edge-direction distribution.

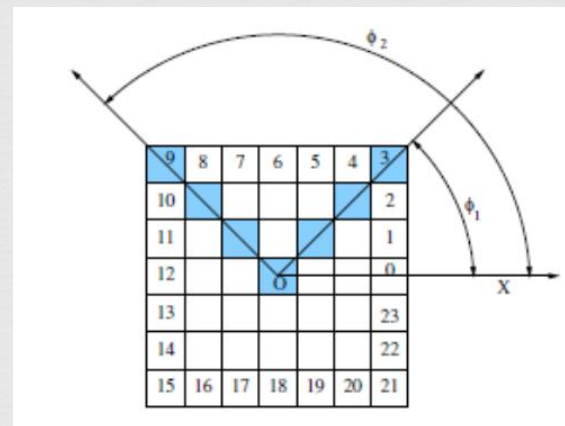
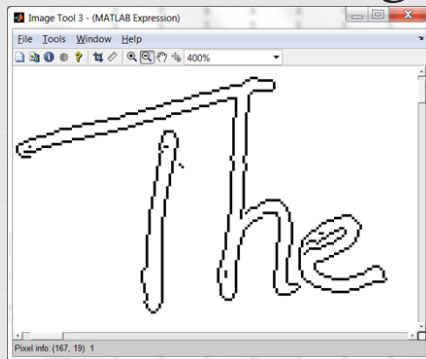


*M. Bulacu, L. Schomaker, and L. Vuurpijl. "Writer identification using edge-based directional features."
In Proceedings of ICDAR 2003, pages 937-941, Edinburgh, UK, 2003

Edge-Hinge Distribution*

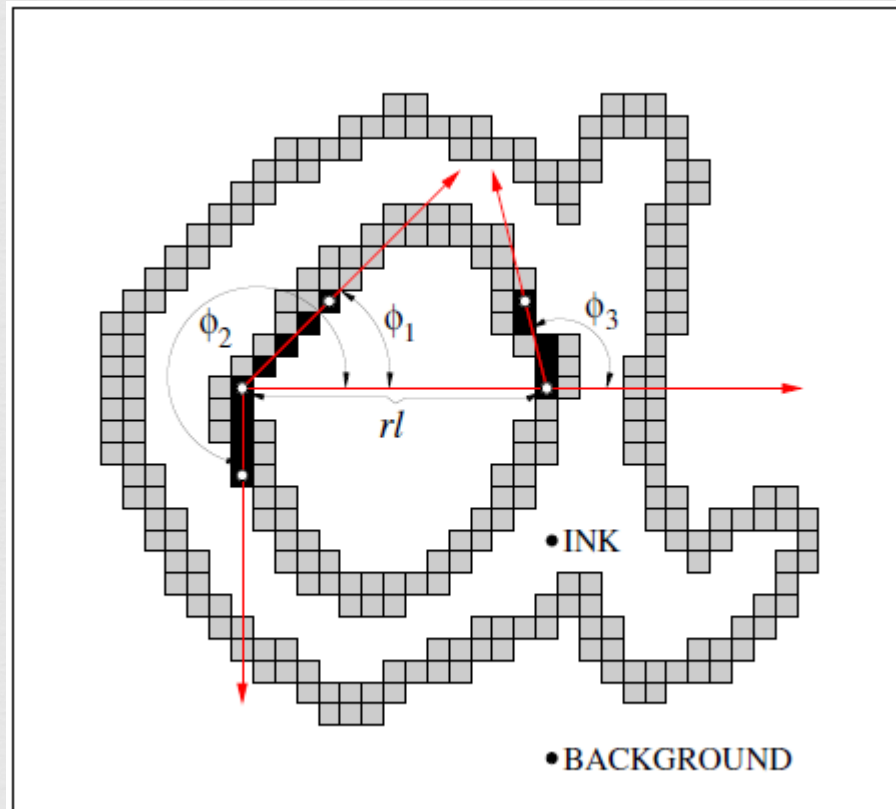


- ✓ Edge detection.
- ✓ Directions of edges ϕ_1 , ϕ_2 with $\phi_2 > \phi_1$.
- ✓ Histogram of Directions.
- ✓ Normalized to a probability distribution $p(\phi_1, \phi_2)$.
- ✓ Nearest Neighbor



Accuracy: 63% for 250 writers.

Edge-Hinge Distribution

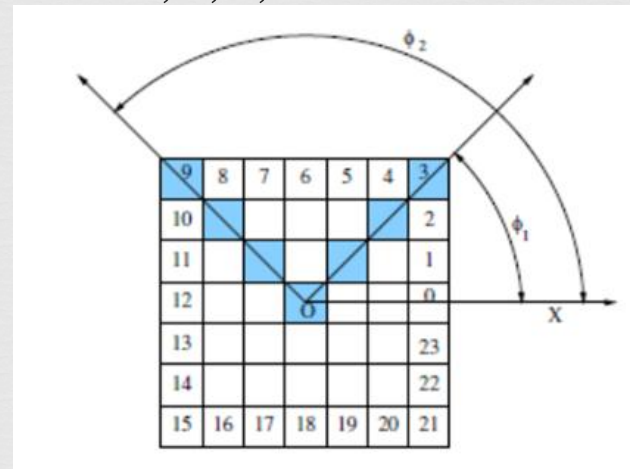


Edge-Hinge Combinations*



- ✓ Improvement of Edge-Hinge Distribution.
- ✓ Similar with Edge-Hinge Distribution.
- ✓ Multiple length edge fragments (windows sizes).
- ✓ Best results on size combinations 3,5,7,9.

Accuracy: 81% for 250 writers.

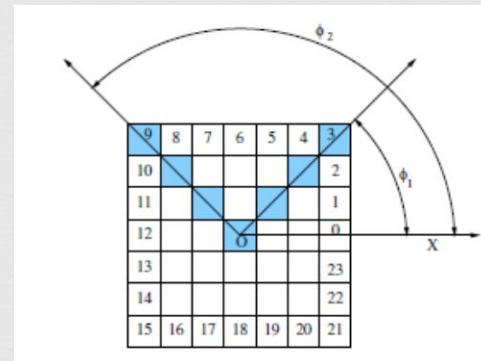
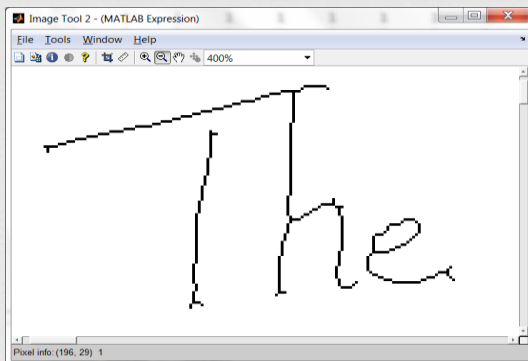


*Van Der Maaten, Laurens, and Eric O. Postma. "Improving automatic writer identification." BNAIC. 2005.

Skeleton-Hinge Distribution



- ✓ Improvement of previous methods.
- ✓ Similar technique.
- ✓ Skeleton information.
- ✓ Hypothesis: All stroke widths should be considered the same
- ✓ Shorter execution time by 35%.

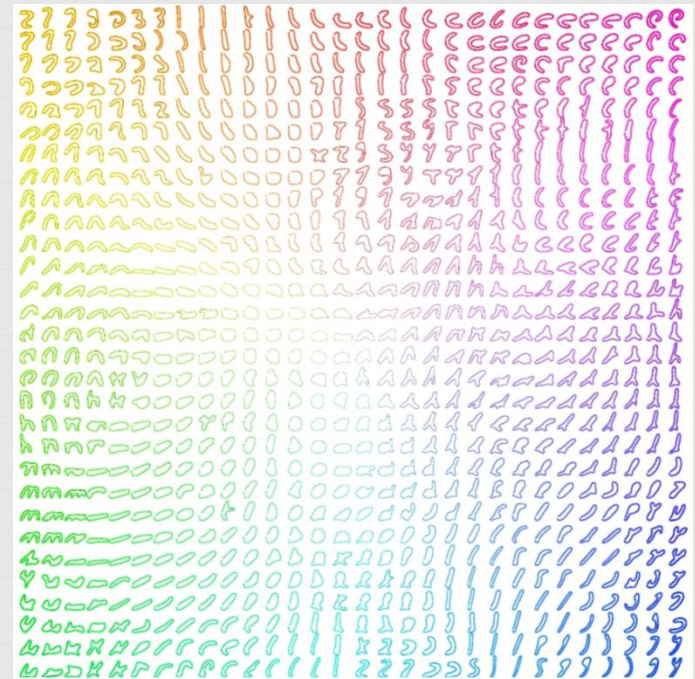
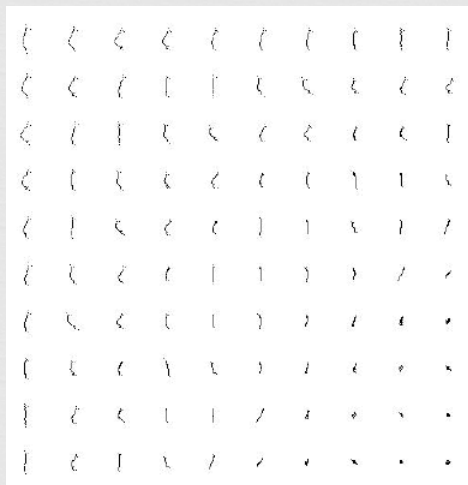


Accuracy: 90,8% for 250 writers.

Codebook of Models of Graphemes*




- ✓ Hypothesis: the writer acts as a stochastic generator of ink-blob shapes, or graphemes.



*Bulacu, Marius, and Lambert Schomaker. "Text-independent writer identification and verification using textural and allographic features." Pattern Analysis and Machine Intelligence, IEEE Transactions on 29.4 (2007): 701-717.

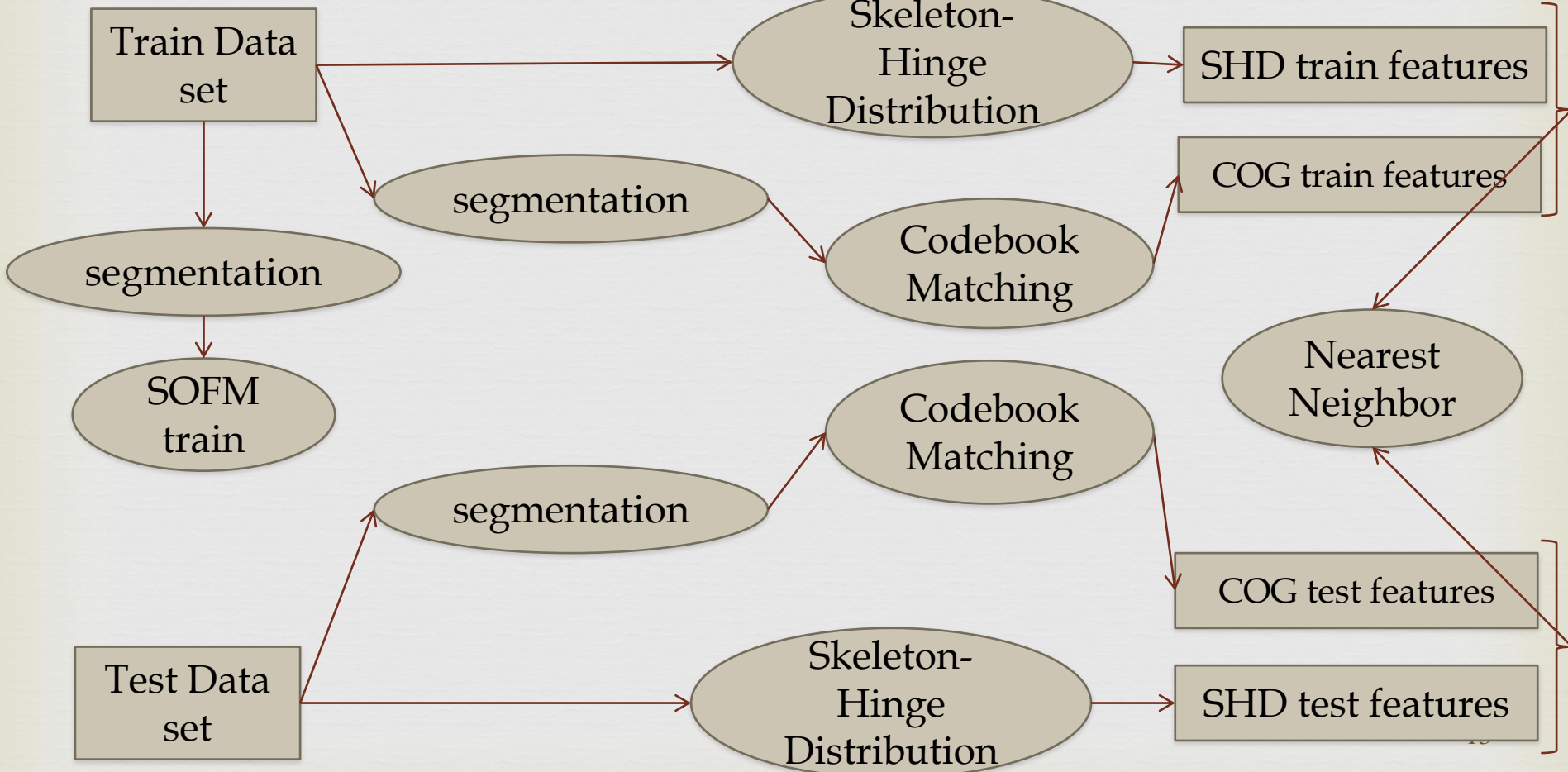
Codebook of Models of Graphemes*



- ✓ Words gets fragmented to graphemes.
- ✓ Grapheme codebook generation using self-organizing feature map (SOFM)
- ✓ Histogram of grapheme models
- ✓ Nearest Neighbor

*Bulacu, Marius, and Lambert Schomaker. "Text-independent writer identification and verification using textural and allographic features." *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 29.4 (2007): 701-717.

System Approach



Data*



- ✓ Firemaker DB
- ✓ 250 writers
- ✓ 4 pages per writer
 - ✓ 1st page: Copied text
 - ✓ 2nd page: Freestyle text
 - ✓ 3rd page: Upper case copied text
 - ✓ 4th page: Copied text with forge attempt
- ✓ Train set : 1st page
- ✓ Test set: 2nd page

*L. Schomaker and L. Vuurpijl. «Forensic writer identification: A benchmark data set and a comparison of two systems [internal report for the Netherlands Forensic Institute].» Technical report, Nijmegen: NICI, 2000.

Skeleton-Hinge Distribution Experimental Results



<u>Fragment Length</u>	<u>Manhattan distance Performance</u>	<u>Euclidian distance Performance</u>	<u>Chi-square distance Performance</u>
<u>3</u>	<u>80%</u>	<u>72%</u>	<u>53.2%</u>
<u>5</u>	<u>89,6%</u>	<u>77,2%</u>	<u>66%</u>
<u>7</u>	<u>90%</u>	<u>81,6%</u>	<u>69,6%</u>
<u>9</u>	<u>88%</u>	<u>85,2%</u>	<u>76%</u>
<u>3, 5</u>	<u>85,2%</u>	<u>75,2%</u>	<u>58,4%</u>
<u>3, 7</u>	<u>85,6%</u>	<u>75,6%</u>	<u>55,2%</u>
<u>3, 9</u>	<u>86%</u>	<u>74,8%</u>	<u>53,2%</u>
<u>5, 7</u>	<u>90%</u>	<u>78,8%</u>	<u>64,4%</u>
<u>5, 9</u>	<u>90.8%</u>	<u>78,8%</u>	<u>67,2%</u>
<u>7, 9</u>	<u>90%</u>	<u>83,2%</u>	<u>73,6%</u>
<u>3, 5, 7</u>	<u>86,8%</u>	<u>76,8%</u>	<u>60%</u>
<u>3, 7, 9</u>	<u>89,6%</u>	<u>76,8%</u>	<u>55,6%</u>
<u>5, 7, 9</u>	<u>90%</u>	<u>79,2%</u>	<u>68,8%</u>
<u>3, 5, 7, 9</u>	<u>89,6%</u>	<u>76,8%</u>	<u>60,4%</u>

Experimental Results



<u>Number of writers</u>	<u>Codebook Size</u>	<u>Manhattan distance Performance</u>	<u>Euclidian distance Performance</u>	<u>Chi-square distance Performance</u>
<u>250</u>	<u>225</u>	<u>95.6%</u>	<u>91.2%</u>	<u>78.8%</u>
<u>150</u>	<u>225</u>	<u>96%</u>	<u>94.7%</u>	<u>86.7%</u>

Comparison



Method	Accuracy
Edge Direction Distribution	35%
Edge-Hinge Distribution	63%
Edge-Hinge Combinations	81%
<u>Skeleton-Hinge Distribution</u>	90.8%
System Approach ¹	96%
Schomaker Approach ^{2*}	97%

1 Codebook size 15 X 15

2 Codebook size 33 X 33

*L. Schomaker, M. Bulacu, and K. Franke. Automatic writer identification using fragmented connected-component contours. In Proceedings of the 9 th IWFHR, pages 185–190, Tokyo, Japan, 2004

Conclusions



- ✓ A single statistical feature achieves high accuracy
- ✓ Our hypothesis proved right.
- ✓ Codebook of graphemes combined with skeleton hinge reached accuracy of 96%.

Questions



Thank you

*«If I have seen further it is by standing on the shoulders
of giants.»*