Assessing Textural Features For Writer Identification on Different Writing Styles and Forgeries

Luiz Eduardo S. Oliveira

Universidade Federal do Paraná, Curitiba, Brazil SticAmSud - Rouen, France - December 2014

Paper published at the International Conference on Pattern Recognition (Bertolini, D., et al.) Stockholm, Sweden, August 24-28, 2014.



Writer Identification

• It concerns the retrieval of handwritten samples from a database using the handwritten sample under study as a graphical query.



Writer Identification

• It concerns the retrieval of handwritten samples from a database using the handwritten sample under study as a graphical query.

He slapped himself in the face and cuffed the sides of his head. Then by degrees the rotating objects slowed, and caring into focus took

- Who wrote this document?
- Several applications such as forensics, digital libraries, and text retrieval.



Introduction

Different approaches for Writer Identification

- Local approach
 - It takes into account specific features of the handwriting and, in general, involves a segmentation algorithm.
 - Local analysis.

Writer 1
 Writer 2
 Writer 3

 'K'
 'M'
 'g'
 'K'
 'M'
 'g'

$$\mathcal{K}$$
 \mathcal{M}
 \mathcal{G}
 \mathcal{K}
 'M'
 'g'

 \mathcal{K}
 \mathcal{M}
 \mathcal{G}
 \mathcal{K}
 \mathcal{M}
 'g'



Introduction

Different approaches for Writer Identification

- Local approach
 - It takes into account specific features of the handwriting and, in general, involves a segmentation algorithm.
 - Local analysis.

- Global approach
 - Tries to avoid segmentation
 - Analysing the look and feel of the writing



Texture Representation

• Create a texture from the handwriting by concatenating all the connected components detected in the binary image [IJDAR 2012].

De Fernando Quintás Zanon Rua Luiz Kirt Wa Kerez, 87 - Ap. 300 Xenapolis, Nova Yolanda 14506-158 Enaldenia BURADOR ALIANA VOID (a) Defensa Q-4ano ZVINNA VOID Maenos-1, Para BUL-Ob Draio e eavo 100 Se orta-060-LUNIA Depakafde thatout ontarienc aursur musicosecoira-1011) A/NE (b)



Texture Representation

• Create a texture from the handwriting by concatenating all the connected components detected in the binary image [IJDAR 2012].

De Fernando Quintas Zanon Rua Luiz Kirt Wa Kerez, 87 - <u>Ap. 300</u> Xenapolis, Nova Yolanda 14506-159 (a) Defens Q: taxo 2: UNALOIO Pacenos 1, Para But Ob Varo e e avo 100 Se Otta OCO-LUNA Depata f de: that out onta vienc aursur musicosecoura 1011 A/DE (b)





Texture Representation

Some observable features in the texture image





- Interesting approach to deal with a large number of classes and few samples per class.
- Dichotomy transformation is used to reduce any pattern classification problem to a 2-class problem.





Dissimilarity Framework

- Each writer contributes with a certain number of references in the feature space
- Positive samples:
 - Compute the difference vectors among the samples of the same writer
- Negative samples:
 - Compute the difference vectors among the samples of different writers selected randomly.



- It uses a disjoint set of writers for training and testing.
- Writers that were not used for training can be identified by the system.
- When a new writer is enrolled into the system, there is no need to retrain the classifier.



- It uses a disjoint set of writers for training and testing.
- Writers that were not used for training can be identified by the system.
- When a new writer is enrolled into the system, there is no need to retrain the classifier.
- Identifying a given writer
 - Extract features from the questioned sample (Q)
 - Extract features from the writer's references (S_i)
 - Compute the dichotomy transformation to create the dissimilarity vectors (*V_i*)
 - Classify V_i and combine the results to produce a decision D
 - Select the writer the maximizes D



- Assess the texture-based writer identification on different handwriting styles and forgeries
- State-of-the-art results were achieved on public databases such as IAM and BFL [ESwA 2013].
- Firemaker database
 - 250 writers
 - Copied text
 - Upper case
 - Forged text (please write as if to impersonate another person)
 - Free (describe a cartoon)



- SVM classifiers
- Two textural descriptors:
 - Local Binary Patterns (LBP^{u2}_{8.2}) [59 components]
 - Local Phase Quantization (LPQ) [256 components]
- Number of writers
 - Training = [20, 100]
 - Testing = [150]
- Number of references: 9 for training and testing.
- Evaluate different writing styles for training.



Results - Different writing styles

Training = 100, Testing 150

Writing Sytle		Descriptors		[3]
Training	Testing	LBP	LPQ	
Сору	Сору	2.0	2.0	3.0
Upper case	Upper case	9.0	6.0	30.0
Free	Free	7.0	7.0	30.0
Mix	Сору	10.0	5.0	-
Mix	U. case	21.0	13.0	-
Mix	Free	24.0	22.0	-



Results - Different writing styles

Training = 100, Testing 150

Writing Sytle		Descriptors		[3]
Training	Testing	LBP	LPQ	
Сору	Сору	2.0	2.0	3.0
Upper case	Upper case	9.0	6.0	30.0
Free	Free	7.0	7.0	30.0
Mix	Сору	10.0	5.0	-
Mix	U. case	21.0	13.0	-
Mix	Free	24.0	22.0	-

Training = 20, Testing 150

Writing Sytle		Descriptors	
Training	Testing	LBP	LPQ
Сору	Сору	4.0	2.0
Upper case	Upper case	21.0	9.0
Free	Free	13.0	11.0
Mix	Сору	5.0	6.0
Mix	U. case	23.0	14.0
Mix	Free	26.0	20.0

Results - Different writing styles

ROC - LPQ-based classifier





Results - Forgeries

Training = 100, Testing 150

Writing Sytle		Descriptors		[3]
Training	Testing	LBP	LPQ	
Сору	Forged	22.0	16.0	-
Upper case	Forged	9.0	9.0	-
Free	Forged	8.0	6.0	-
Mix	Forged	16.0	14.0	50.0



Results - Forgeries

Training = 100, Testing 150

Writing Sytle		Descriptors		[3]
Training	Testing	LBP	LPQ	
Сору	Forged	22.0	16.0	-
Upper case	Forged	9.0	9.0	-
Free	Forged	8.0	6.0	-
Mix	Forged	16.0	14.0	50.0

Training = 20, Testing 150

Writing Sytle		Descriptors	
Training	Testing	LBP	LPQ
Сору	Forged	22.0	22.0
Upper case	Forged	16.0	14.0
Free	Forged	12.0	10.0
Mix	Forged	24.0	28.0



Results - Forgeries

LPQ-based classifier



Example of a forged text

Bob, David on sexy Kantippe sparen postecques van de landen Egypte, Japan, algerije, de USA, Holland, Mtalic Grickenland (a)

Nog dezelfde avond reden ze na han wienden Chris Emile, Jan, Trene, en Henk, nadat ze hun



⁽b)

- Discussed the use of textural descriptors for writer identification on different writing styles and forgeries.
- LPQ-based classifier achieved encouraging results and compare favourably to other methods.
- The number of writers used for training becomes important to deal with forgeries.
 - Writer selection



Thank you for your attention !

