Innovative Methods for Non-Destructive Inspection of Handwritten Documents

Eleonora Breci, Luca Guarnera, Sebastiano Battiato Department of Mathematics and Computer Science, University of Catania, Italy brecieleonora@gmail.com, luca.guarnera@unict.it, sebastiano.battiato@unict.it



Abstract

Handwritten document analysis is an area of forensic science, with the goal of establishing authorship of documents through examination of inherent characteristics. Law enforcement agencies use standard protocols based on manual processing of handwritten documents. This method is time-consuming, is often subjective in its evaluation, and is not replicable. To overcome these limitations, in this paper we present a framework capable of extracting and analyzing intrinsic measures of manuscript documents related to text line heights, space between words, and character sizes using image processing and deep learning techniques. The final feature vector for each document involved consists of the mean (η) and standard deviation (σ) for every type of measure collected. By quantifying the Euclidean distance between the feature vectors of the documents to be compared, authorship can be discerned. Our study pioneered the comparison between traditionally handwritten documents and those produced with digital tools (e.g., tablets). Experimental results demonstrate the ability of our method to objectively determine authorship in different writing media, outperforming the state of the art

Proposed pipeline

(a) Comparison between documents. (b) Feature extraction module: 1st module): lines of text and words are automatically detected. Then, one or more characters (templates) chosen by the expert can be searched within the document. 2^{st} module): the feature vector is defined as the means (η) and standard deviations (σ) of all

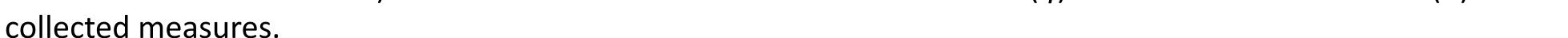
Results

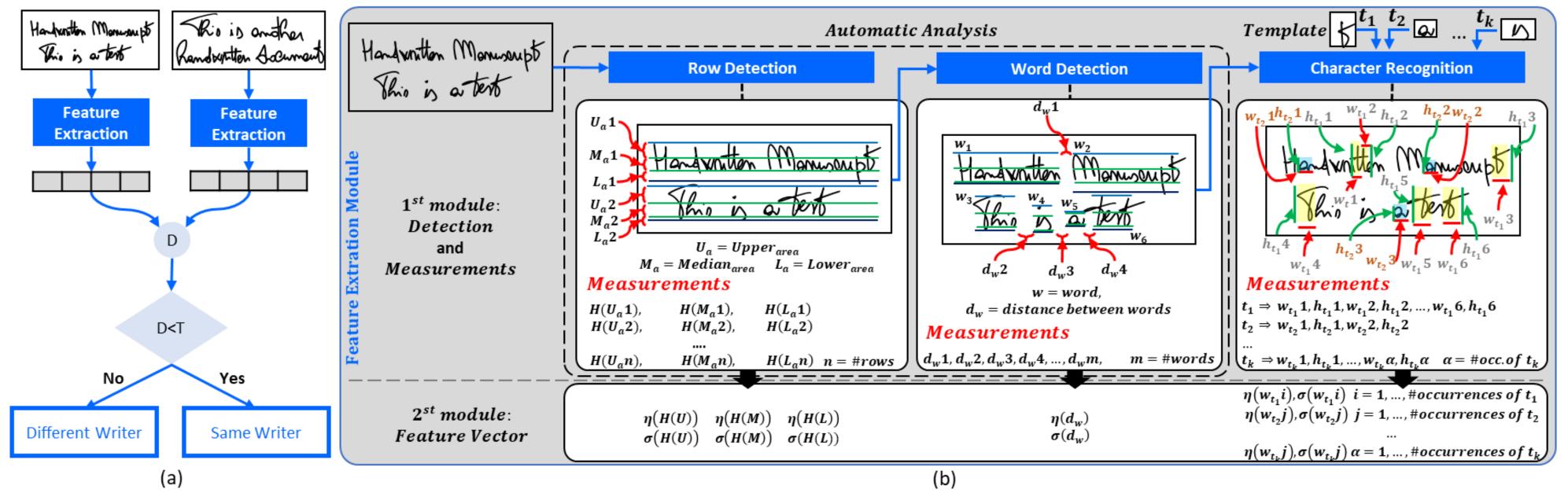
(1) Testing with state-of-the-art datasets

You have helded my low. You used to stir	A Political Correspondent writes: Ban the
my impirator. Non you don't even sturmy	sents demanstrations througed the Central
curiosity. You is mply produce no elfect.	colleg af the secure of camero last
I loved you because you were morvellow,	night and formed a quere delaring
because you had genies and intellect,	for more know 200 yevers autoine
eccuse you realised the dreams of gread	in our Palace Yard, Alaut 2000
vets and grave shape and substance to	abbied M. 7. s and harangued men
e shadows of art. You have thrown I	an disarmanent.
away. Yo one shallow and stypiol.	and only in and it.
	Should the Herald publish such advertisements?
Werd ich eum Augenbliche agen:	This is a difficult question. It would obviously
Venale dach! du bost as action!	•
	be wrong to refuse all political advertisements with
Dorn magel du nich in Fersch schlagen,	which we disagree. When an advertisement
Dann will ich gen en Gande gehn!	contains statements whose factual touth is
	doubtful, or where the total content would be
Dan ope die Tadle date challe.	
Non may die Todlenglahe schalle,	deeply repugnant to an readers, it is right to
Dan mag die Todlenglache schalle, Dan biel de demes Diendes frey,	deeply repugnant to an reader, it is right to

٨	٨	Poli	ical	Corneapon	responselut w		Ban - the .	
		82	al 9	- data	1.	a hu	Gu	was

(b)





Proposed Approach

Technique 1: Text Line and Word Detection

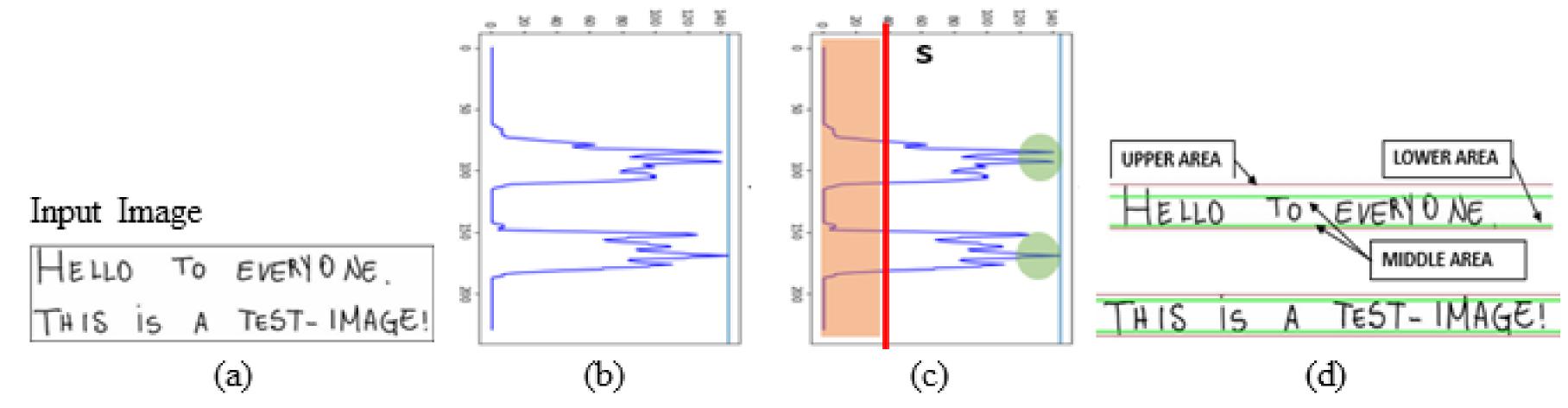
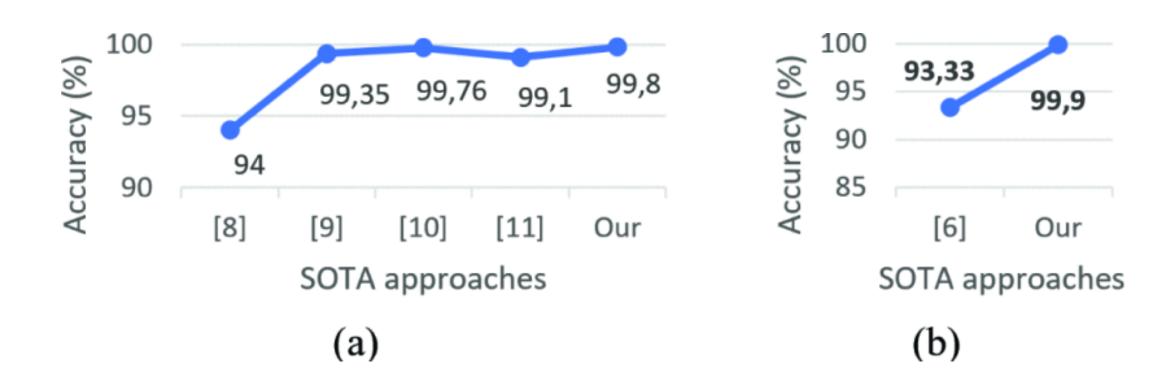
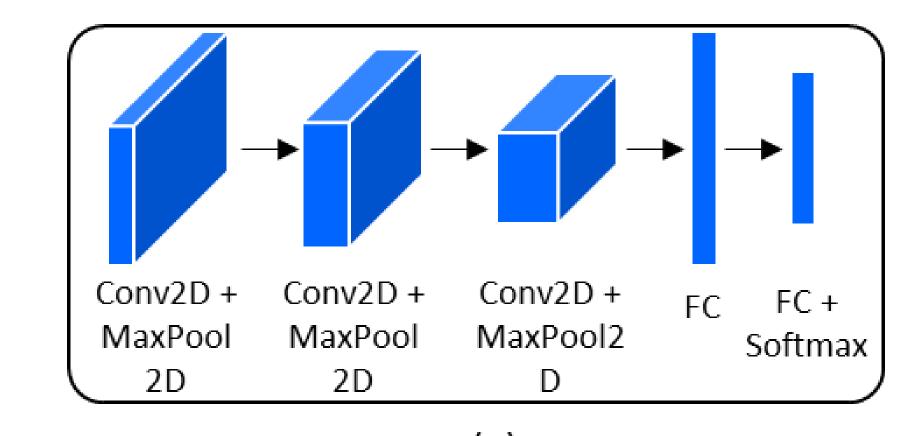


Fig. Examples of (a) CVL and (b) CSAFE digitized texts



Technique 2: Character Recognition



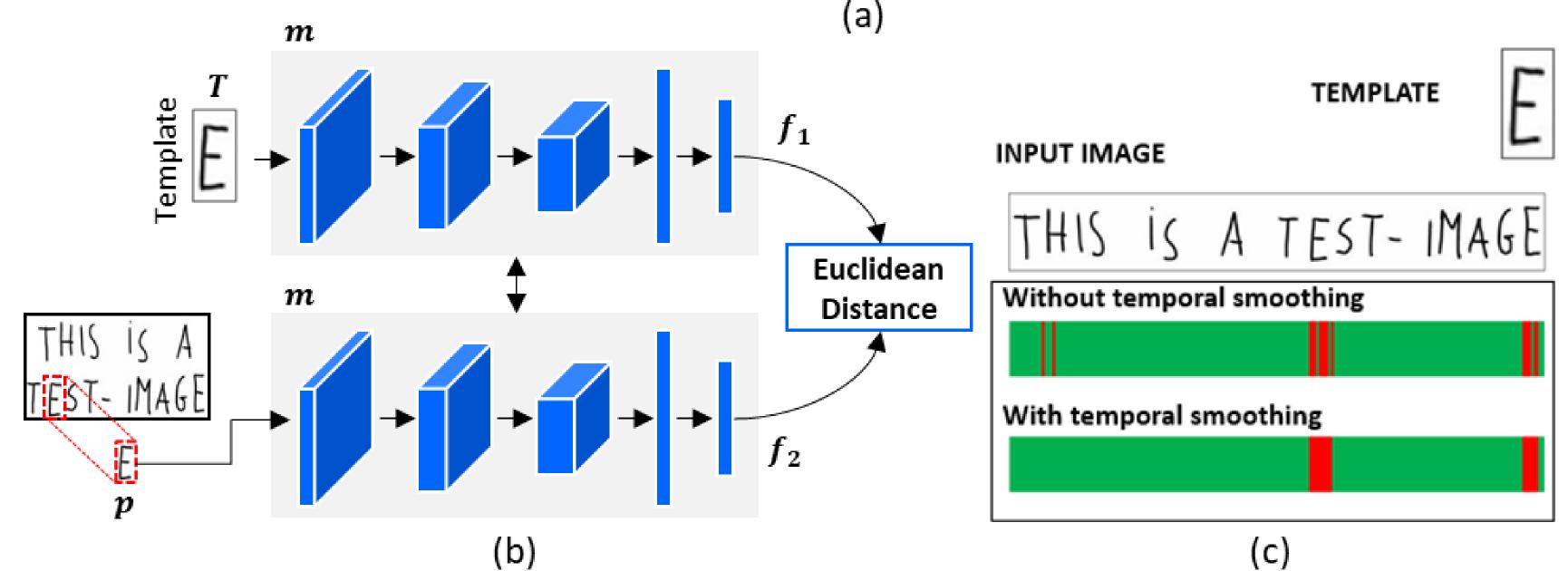
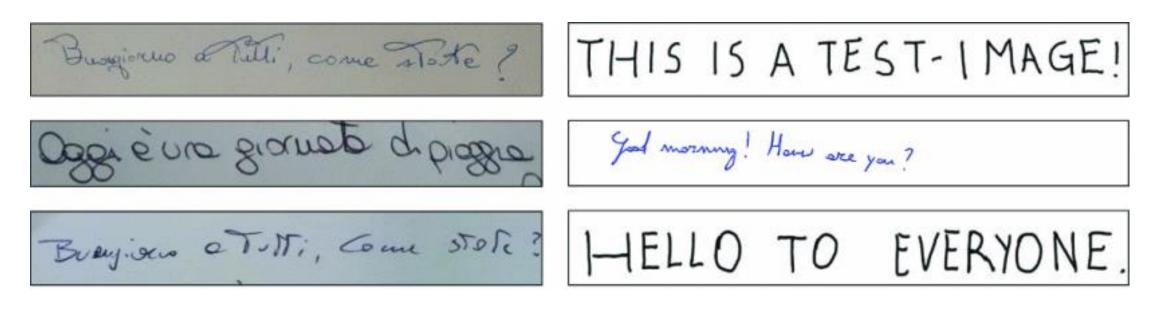


Fig. Comparison with SOTA approaches, by using CVL (a) and CVL + CSAFE datasets (b).

2) Preliminary results with a customized dataset:

- Images acquired from the classic "pen and paper" approach.
- Images acquired using tablets.

(a)



All documents were compared with each other, including handwritten and digital documents, resulting in a classification accuracy of 96%.

Web Page



https://iplab.dmi.unict.it/mfs/forensic-handwriting-analysis/innovative-methods-2023/

References

- 1. Luca Guarnera, et al., "GRAPHJ: A Forensics Tool for Handwriting Analysis," in Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part II 19. Springer, 2017, pp. 591–601
- 2. Eleonora Breci, et al., "A Novel Dataset for Non-Destructive Inspection of Handwritten Documents," in IS&T International Symposium on Electronic Imaging: Media Watermarking, Security, and Forensics, MWSF, 2024
- 3. Luca Guarnera, et al., "Forensic Analysis of Handwritten Documents with GRAPHJ," Journal of Electronic Imaging, vol. 27, no. 5, pp. 051230-051230,2018.
- 4. "A-z handwritten alphabets kaggle. format. .CSV IN https://www.kaggle.com/sachinpatel21/azhandwrittenalphabets-in-csvformat, 2023.
- 5. Sheng He and Lambert Schomaker, "Fragnet: Writer Identification Using Deep Fragment Networks," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3013–3022, 2020.
- 6. Amy M Crawford, et al., "A Statistical Approach to Aid Examiners in the Forensic Analysis of Handwriting," Journal of Forensic Sciences, 2023.
- 7. Abdelillah Semma, et al., "Writer Identification Using Deep Learning with Fast Keypoints and Harris Corner Detector," Expert Systems with Applications, vol. 184, pp. 115473, 2021.
- 8. Akram Bennour, et al., "Handwriting Based Writer Recognition Using Implicit Shape Codebook," Forensic Science International, vol. 301, pp. 91–100, 2019.
- Parveen Kumar and Ambalika Sharma, "Segmentation- free Writer Identification Based on Convolutional Neu- ral Network," Computers & Electrical Engineering, vol. 85, pp. 106707, 2020
- 10. Songxuan Lai, et al., "Encod- ing Pathlet and SIFT Features with Bagged VLAD for Historical Writer Identification," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3553– 3566, 2020.
- 11. Sheng He and Lambert Schomaker, "Fragnet: Writer Identification Using Deep Fragment Networks," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 3013–3022, 2020