REAL-TIME DOCUMENT DETECTION IN SMARTPHONE VIDEOS

A Mathematical Morphology Approach

Élodie Puybareau and Thierry Géraud

EPITA Research and Development Laboratory (LRDE), France
firstname.lastname@lrde.epita.fr

At a glance

Problem:

- Real-time document detection in smartphone videos is challenging [3]

Why our approach is interesting:

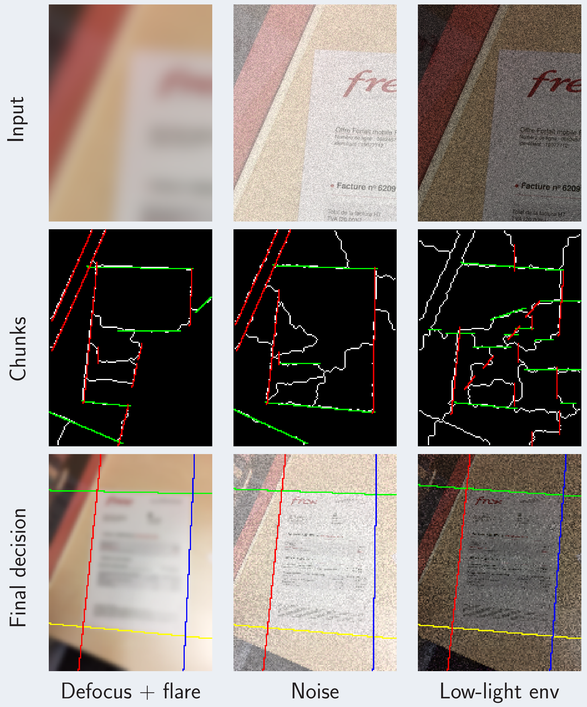
- based on classical mathematical morphology operators [6]
- no *a priori* on documents in images
- light enough to be run on smartphones

Conclusion:

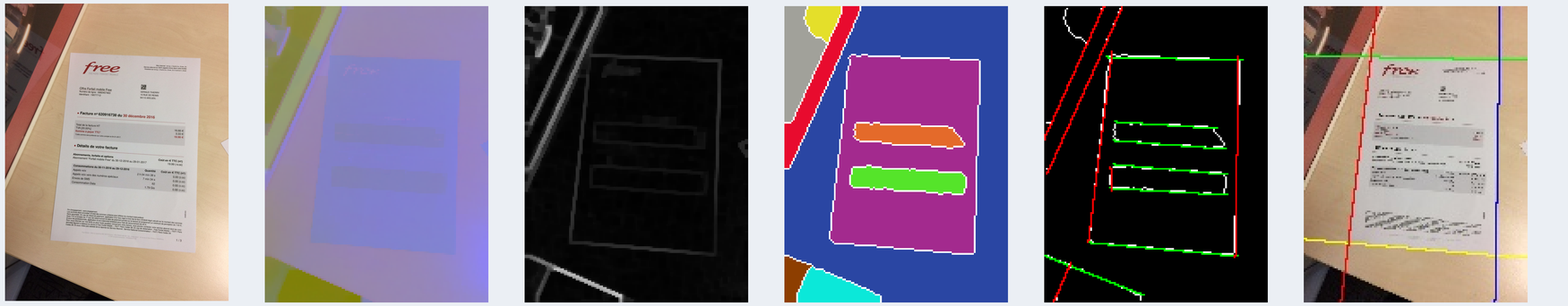
our method is

- **fast** (0.04 s per frame)
- **robust** (to many defects: noise, defocus, moves, low-light...)
- and **effective** (Jaccard coefficient of 0.9 on SmartDoc 2015 [5])

Robustness of our method



Step by step



Description of the method

- Pre-processing:
  - ⇝ reduction of each frame to 180 × 100 px, and conversion to La^*b^* space
  - ⇝ regularization with a morphological closing on L, and an erosion on a
- Segmentation of the image into regions:
  - ⇝ morphological thick gradient on each component of La^*b^*, and summed up
  - ⇝ morphological closing to remove non-significant regional minima
  - ⇝ morphological watershed transform ⇒ a collection of basins
- Extraction of line chunks from region contours:
  - ⇝ Hough transform on the binary watershed line image,
  - ⇝ post-processing: cut lines into chunks, and remove redundant chunks
- Finding the document boundaries:
  - ⇝ classification of chunks (top, bottom, left, right)
  - ⇝ pairing of compatible chunks (left-top for example)
  - ⇝ finding the best path (i.e., left-top + top-right + etc.) using an energy criterion

Quantitative results

Method	set#1	set#2	set#3	set#4	runtime
Xu et al. [2]	0.997	0.987	0.999	0.994	>1min
LRDE SmartDoc	0.987	0.977	0.989	0.984	>1min
Leal et al. [1] (best)	0.961	0.944	0.965	0.930	0.43s
SmartDoc average [5]	*0.946*	*0.903*	*0.938*	*0.812*	*?*
Leal et al. [1] (fastest)	0.921	0.849	0.909	0.840	0.10s
Our	0.905	0.936	0.859	0.903	**0.04s**

Some qualitative results

Selected bibliography

[1] L.R. Leal and B.L. Bezerra, "**Smartphone camera document detection via geodesic object proposals**," in *IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, pp. 1–6, 2016.

[2] Y. Xu, E. Carlinet, T. Géraud, and L. Najman, "**Hierarchical segmentation using tree-based shape spaces**," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 457–469, 2017.

[3] J. Liang, D. Doermann, and H. Li, "**Camera-based analysis of text and documents: A survey**," *International Journal on Document Analysis and Recognition*, vol. 7, no. 2, pp. 84–104, 2005.

[4] M. Ôn Vũ Ngọc, J. Fabrizio, and T. Géraud, "**Saliency-based detection of identy documents captured by smartphones**," in *IAPR International Workshop on Document Analysis Systems (DAS)*, pp. 387–392, 2018.

[5] J. Burie et al., "**ICDAR 2015 competition on smartphone document capture and OCR (SmartDoc)**," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1161–1165, 2015.

[6] L. Najman and H. Talbot, Eds., "**Mathematical Morphology—From Theory to Applications**," ISTE Ltd and John Wiley & Sons Inc, 2010.