Spatial and Color Spaces Combination for Natural Scene Text Extraction

POLYTECH.MONS

Céline Mancas-Thillou and Bernard Gosselin

TCTS Lab, Faculté Polytechnique de Mons (FPMs), Belgium

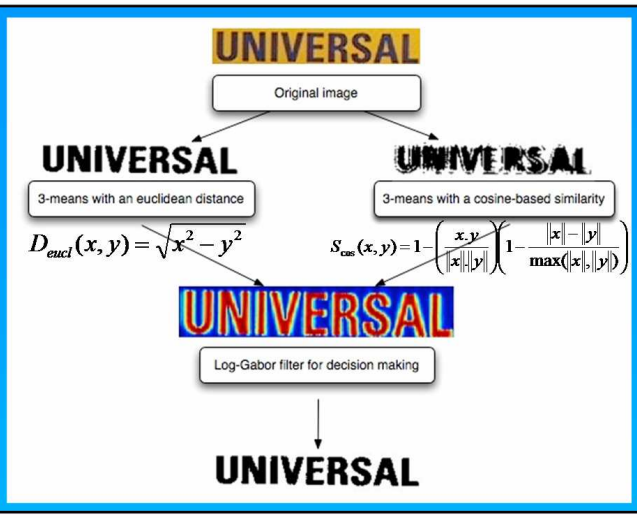
Context

- *Context* – Possibly Low resolution Natural Scene Camera-based Images containing text
- *Our goal* – Text extraction from background to fulfill existing methods and off-the-shelf OCR drawbacks and to handle color variations
- *Main idea* – To circumvent constrained definitions of existing color spaces by combining several clustering metrics inside the RGB color space

Samples of the ICDAR 2003 Database

Method

Brief Overview



For each Clustering

- 3-means clustering on a reduced number of clusters

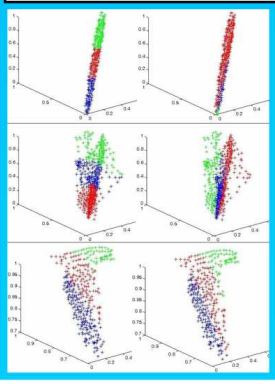
  -> 3 clusters : BACKGROUND, TEXT, EITHER NOISE OR TEXT

- Possibility of combination between two of the three clusters
- Introduction of a new regularity measure : $M=\sum_{i=1}^{N}\left|area_i-\frac{1}{N}\sum_{i=1}^{N}area_i\right|$

  -> MAIN CLUSTER = SMALLEST *M*, MERGING OF CLUSTERS IF *M* DECREASES

Complementarity of Both Metrics :

Euclidean distance and Cosine-based Similarity



- Hue integration thanks to cosine-based similarity inside RGB
- Simultaneous Magnitude and Directional Processing in the same color space
- On a huge database : Deucl performs better in 56% images while Scos in 12%

  -> HOWEVER, ONLY 32% OVERLAP BETWEEN BOTH METRICS
- Different ways of clustering : image on the left
- Different images handling : image on top-right

But How to Choose the Best Clustering ?

-> By using Log-Gabor filters :

- Good for Natural Scene Images
- Add simultaneous information of intensity and spatiality
- Bring additional accuracy on top of color clustering
- Also used in Character Segmentation

-> Equations of Log-Gabor filters in polar coordinates :

$$H(f,\theta)=H_f * H_\theta=\exp\left\{\frac{-[\ln(f/f_0)]^2}{2[\ln(\sigma_f/f_0)]^2}\right\}*\exp\left\{\frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2}\right\}$$

$$\Delta\Omega=2\sigma_\theta\sqrt{2\ln 2}$$

$$B=2\sqrt{2/\ln 2}*\left|\ln(\sigma_f/f_0)\right|$$

->$\Delta\Omega=\Pi/2$, $B=0.3$ octaves, f_0 = Character's thickness

-> Computation of average values of pixels to choose the best segmentation :

BEST CLUSTERING = HIGHEST AVERAGE

Results

- Improvement of 6.3% by combining both metrics compared to the Euclidean distance alone with the public ICDAR 2003 database
- Efficiency of Log-Gabor filters for decision making is improvement of 19.9%.
  - Silhouette information of clusters, indicating how well clusters are separated and dense, performs well in 77.7% images
  - Log-Gabor filtering performs well in 93.2% images

Conclusion and Future Work

Some points already done …

- Combination of two complementary metrics : the Eucldiean distance and a Cosine-based similarity to handle color variations of natural scene images
- Choice of the best clustering by using a dedicated tool for natural scene images : Log-Gabor filtering
- Combination of color, intensity and spatial information through this selective metric clustering called SMC !

… but highlight of the main drawback

- Processing of embossed characters :
  - Use of a priori information and robust OCR

Acknowledgements

This project called SYPOLE is funded by Ministère de la Région wallonne, Belgium.

For this PhD end, thank you to all colleagues!

E-mail: celine.thillou@tcts.fpms.ac.be

FACULTÉ POLYTECHNIQUE DE MONS

Funded by :

ACADÉMIE UNIVERSITAIRE WALLONIE-BRUXELLES