# Content-based document signature: ongoing work

Anh Thu Phan Ho, Petra-Krämer Gomez and Mickaël Coustaty

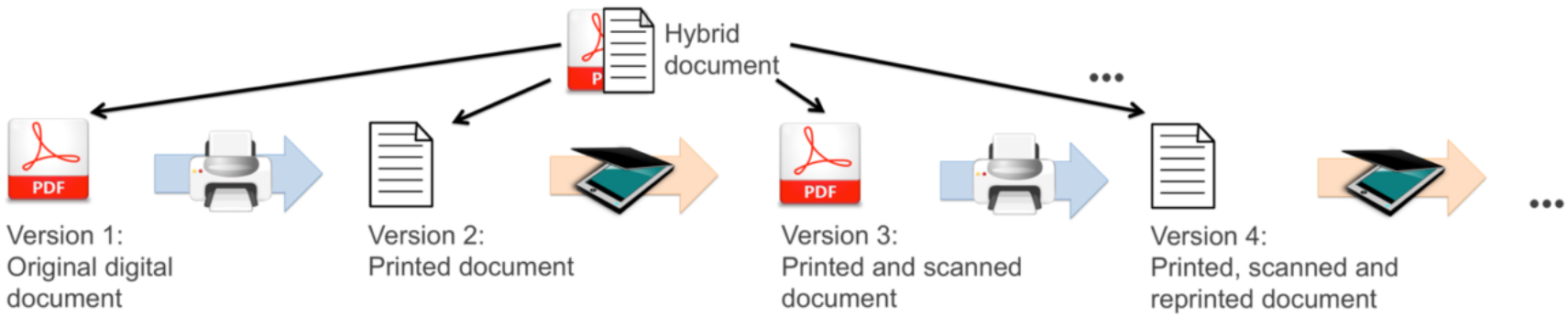
L3i laboratory, University of La Rochelle

GDR Sécurité Informatique réunion 31 Mai

# Outline

# Motivation

- Counterfeit documents are created easily in paper and digital formats
- Paper document security: watermarking and fingerprinting [1],[2]
- Digital document security: electronic signature [3]
- Goal: study hybrid document security by hashing its content



# Outline

# SHADES project

- SHADES: Semantic Hash for Advanced Document Electronic Signature
  - Securing hybrid documents by hashing the document's content
- Partners: researchers from computer science and law domains

# Scheme for SHADES

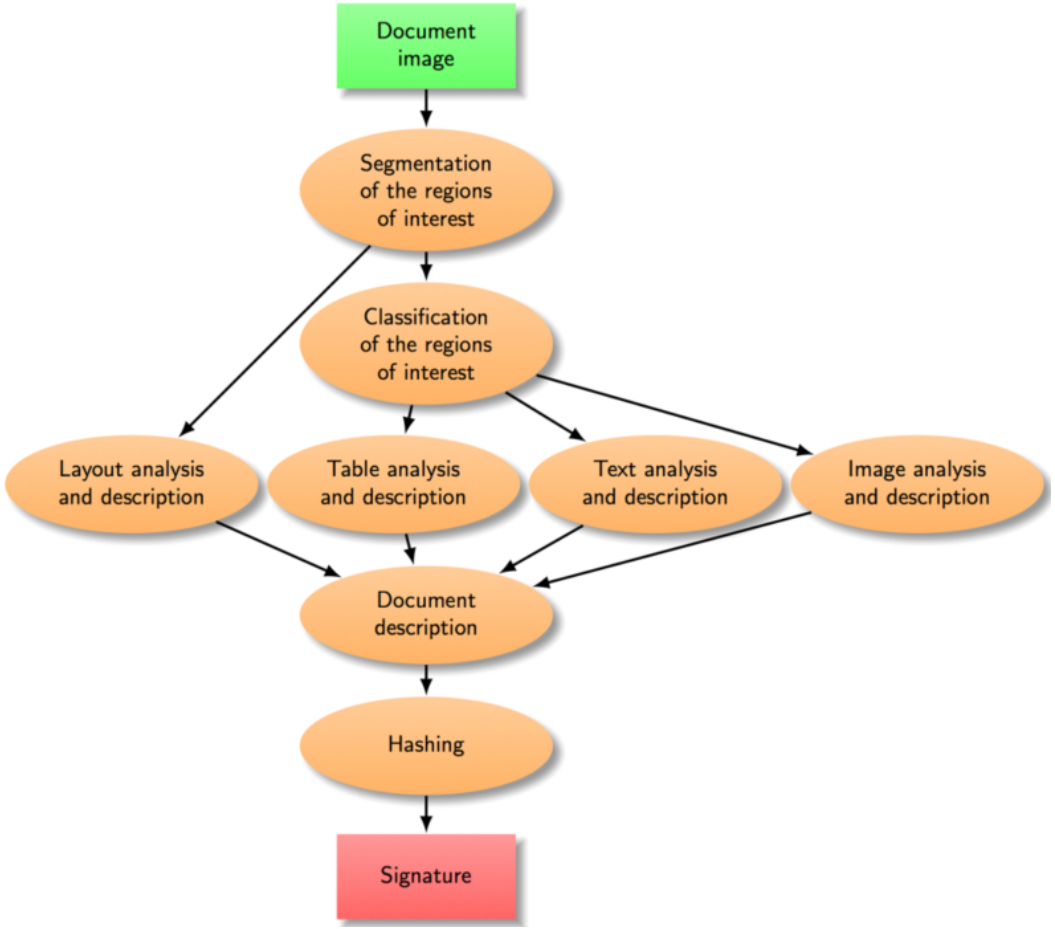
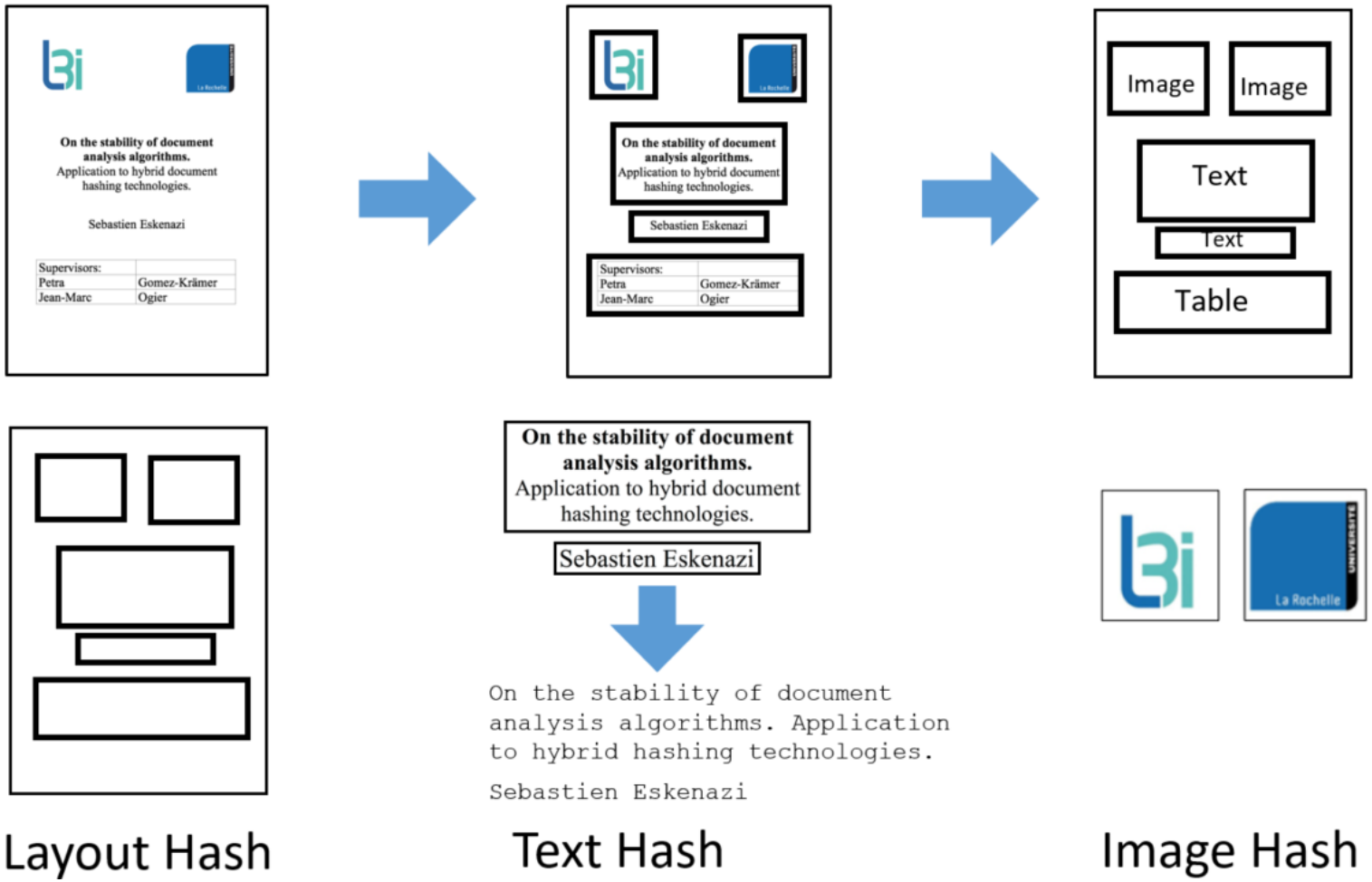



Figure: Algorithm for content-based hashing [Eskenazi, Gomez-Krämer, Ogier 2015]

# Example

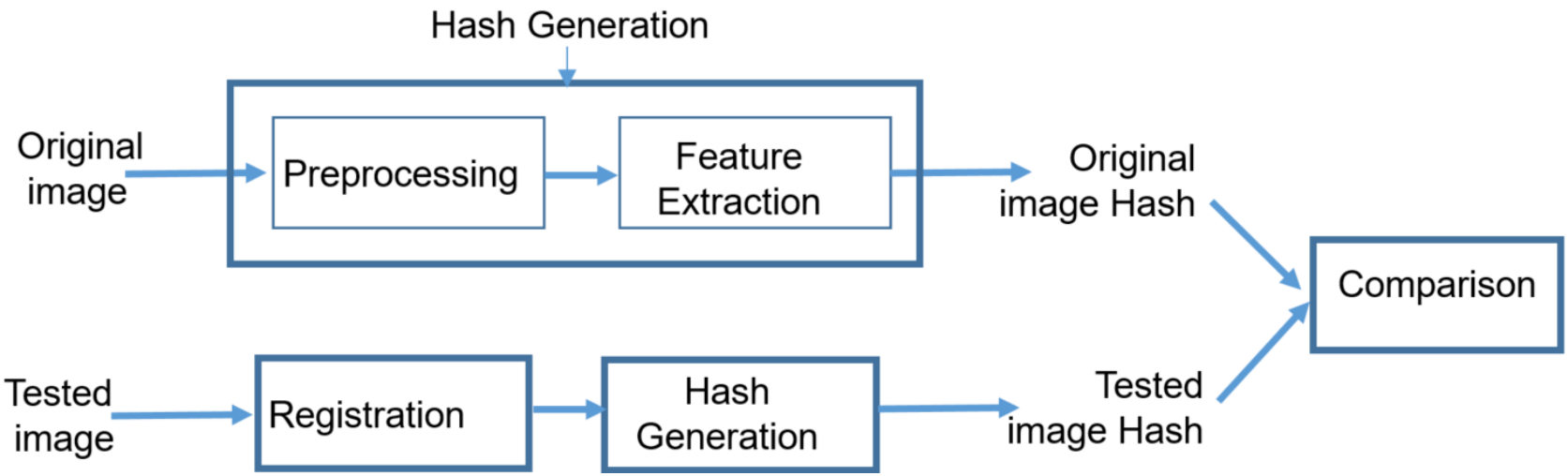

On the stability of document analysis algorithms.
Application to hybrid document hashing technologies.
Sebastien Eskenazi
Supervisors:
Petra Gomez-Krämer
Jean-Marc Ogier
Image
Image
Text
Text
Table
On the stability of document analysis algorithms. Application to hybrid document hashing technologies.
Sebastien Eskenazi
On the stability of document analysis algorithms. Application to hybrid hashing technologies.
Sebastien Eskenazi
Layout Hash
Text Hash
Image Hash

# Outline

# Perceptual Image Hash (PIH)

- PIH is a compact representation used to verify the integrity of an image
- Desirable properties of PIH
  - Unpredictability
  - Robustness against Content Preserving Operations
  - Discrimination to Content Changeing Operations
  - Compactness
- Two types of PIH
  - Unkeyed PIH
  - Keyed PIH

# ASYCHA [Eskenazi, Bodin, Gomez-Krämer, Ogier 2017]




Registration: reduce rotation and scale variations caused by print and scan, need second order moments and resolutions of two images

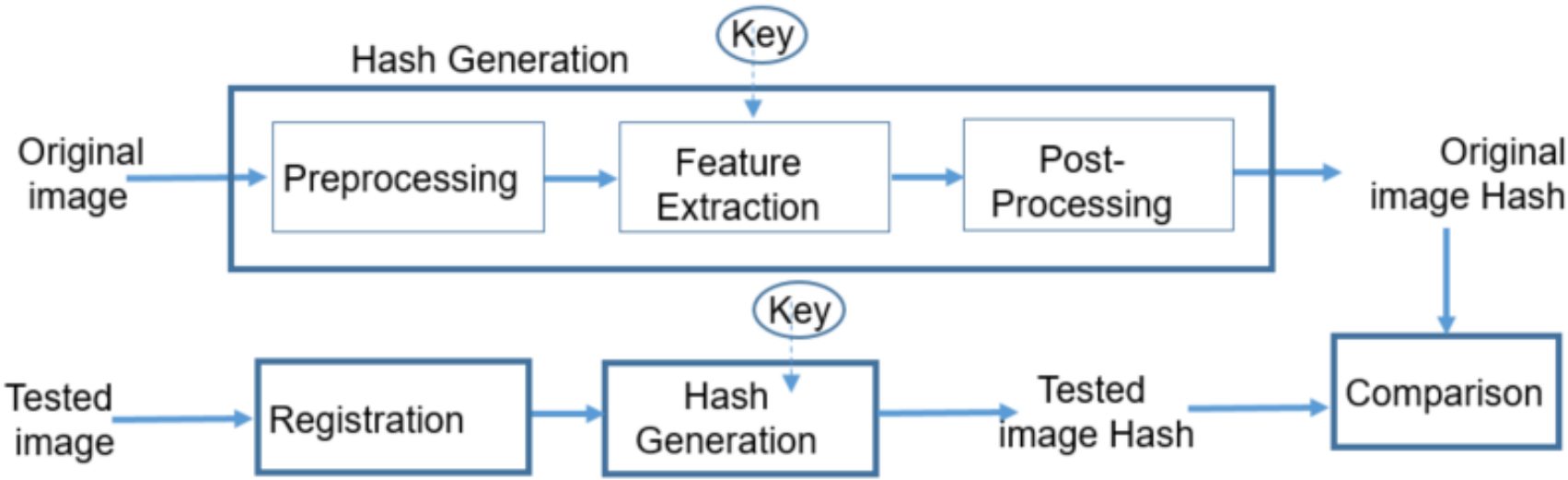
# ASYCHA-Hash Generation

- Preprocessing:
  - Inversion: want to compare the foreground
  - Down-sampling: remove noise and small color variations caused by P&S
  - Indexation on 32 colors: remove colorimetric noise caused by P&S
  - Normalization: compensate for illumination variation caused by P&S
- Feature Extraction:
  - Resolution produced by the scanner
  - Second order moments
  - Index matrix and its color mapping table

# Improvement

- Advantages:
  - High stability w.r.t. P&S
- Improvement:
  - Security: keyed perceptual image hash
  - Hash computation
  - Digest compactness

# Keyed Image Hashing



# Ongoing work

- Keep the same features as those of ASYCHA algorithm
- Add a vector of pseudo random Key
- Compute and compress or quantizatize a hash
- Analyze the security and robustness of this hash
  - Compute the differential entropy of the hash function
  - Use ROC curves to evaluate the robustness and discrimination

# Example

- Features $x_1, x_2, ...x_n$
- Key $K = K_1K_2...K_n$ , where $K_i$ is a pseudorandom variable that is normally distributed with mean $\mu$ and variance $\sigma^2$ [4]
- $i$th hash value $h_i = k_i x_i$
- Differential entropy of $h_i$

$$E(h_i) = \mu x_i$$

$$Var(h_i) = \sigma^2 x_i^2$$

$$\mathfrak{N}(h_i) = \int_{-\infty}^{\infty} h_i \log_2 \frac{1}{h_i} dk = \frac{1}{2}\left[\log_2\left(2\pi e \sigma^2 x_i^2\right) + 1\right]$$

# References


- [1] Aravind et al.Signature-embedding in printed documents for security and forensic applications.Proc. SPIE 5306, Security, Steganography, and Watermarking of Multimedia Contents VI, (22 June 2004)
- [2] Ali al Haj et al. Copyright protection of e-government document images using digital watermarking,Information Management (ICIM), 2017 3rd International Conference on
- [3]Andrew David McCabe Thomas H. Gonser.Systems and methods for distributed electronic signature documents.US Patent US8949706B2, 2015
- [4]Swaminathan, A., Mao, Y., Wu, M.: Robust and secure image hashing. IEEE Transactions on Information Forensics and Security 1(2) (June 2006)

# Thank you for your attention