Mobile visual object identification: from SIFT-BoF-RANSAC to SketchPrint

S. Voloshynovskiy, M. Diephuis, T. Holotyak

University of Geneva
Switzerland

February 11, 2015

Overview

1. Problem under consideration
2. Exisiting technologies and their restrictions
3. Proposed solution: SketchPrint

Problem under consideration (1)

Goal

to develop efficient methods for the identification and security of physical objects based on images acquired from mobile phones

- **Identification**: to establish a type of the object in the group $w \in \{1, \cdots, M\}$ (discover functionalities, augmented reality, 3rd screen, etc.)
- **Security**: to verify the authenticity of object (anti-counterfeiting)

Problem under consideration (1)

Goal

to develop efficient methods for the identification and security of physical objects based on images acquired from mobile phones

- **Identification**: to establish a type of the object in the group $w \in \{1, \cdots, M\}$ (discover functionalities, augmented reality, 3rd screen, etc.)
- **Security**: to verify the authenticity of object (anti-counterfeiting)

Targeted physical objects

- Packaging (pharma, cosmetics, ...),
- Watches (both metal and plastic)
- Electronics (molding)
- Printed documents (incl. text docs, certificates, ID docs, ...)

Remark: no added or embedded features

Problem under consideration (2)

Product identification on mobile phones

Problem under consideration (2)

Product identification on mobile phones

Once identified

- Connect to services: buy, find similar, find on map, check for promotions, check suitability (ingredients, dosage, ...)
- Verify the authenticity: authentic/fake
- Inform brand owners: market study, fake detection ...

Problem under consideration (3)

Particularities of objects

- Very heterogeneous visual content (packages, watches, labels, text docs, microstructures/textures....)
- Similar visual appearance within the same class: many objects look very similar (only small differences)
- Visual features: not very rich

Problem under consideration (3)

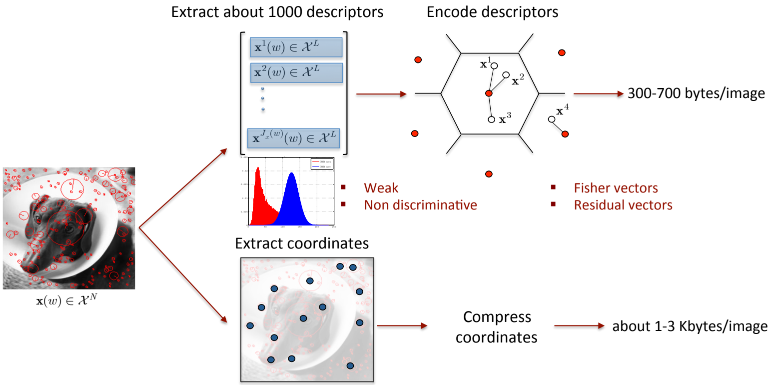
Particularities of objects

- Very heterogeneous visual content (packages, watches, labels, text docs, microstructures/textures....)
- Similar visual appearance within the same class: many objects look very similar (only small differences)
- Visual features: not very rich

Why not digital watermarking?

- All objects should be watermarked: invasive and back-comparability
- Not all objects can be watermarked (watches, etc...)
- Recent theoretical study indicates that visual identification systems are superior to watermarking in terms of identification rate. [Farhadzadeh, Willems, Voloshynovskiy, ISIT2015]

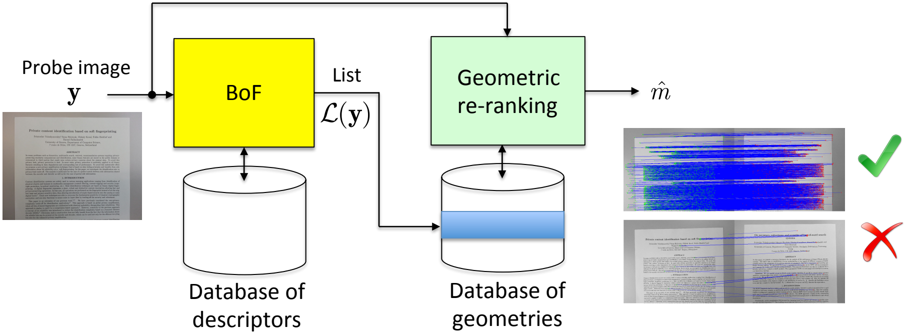
Our goal: identification based on non-invasive technology



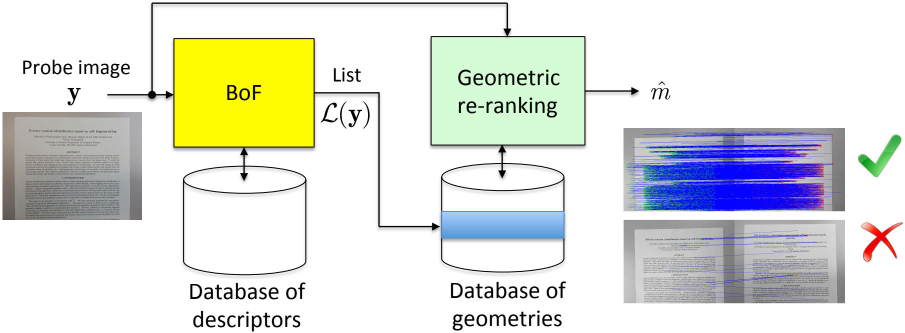
Observations:

- Local features are not very discriminative and quite weak
- Main gain comes from fusion of multiple weak features assuming that some of them will survive ⇒ huge redundancy
- Very complex encoding methods are used to compress this redundancy
- Geometric re-ranking is needed for fine pruning

Exisiting technologies: SIFT+BOF+RANSAC (2)

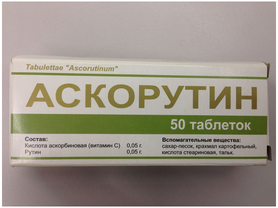


Exisiting technologies: SIFT+BOF+RANSAC (2)



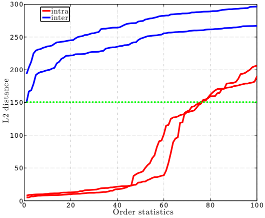
- If BoF fails to produce a short list, then the identification is based only on geometric re-ranking $\Rightarrow$ huge complexity

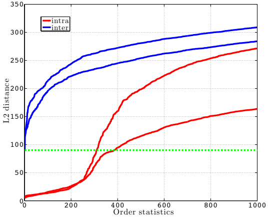
Example of SIFT: real packages (BoF without geometric re-ranking)



100 SIFTs enrolled

1000 SIFTs enrolled





Example of SIFT: text documents (BoF without geometric re-ranking)

Private content identification based on soft fingerprinting

Sviatoslav Voloshynovskiy*, Taras Holotyak, Oleksiy Koval, Fokko Beekhof and Farzad Farhadzadeh
University of Geneva, Department of Computer Science,
7 route de Drize, CH 1227, Geneva, Switzerland

ABSTRACT

In many problems such as biometrics, multimedia search, retrieval, recommendation systems requiring privacy-preserving similarity computations and identification, some binary features are stored in the public domain or outsourced to third parties that might raise certain privacy concerns about the original data. To avoid this privacy leak, privacy protection is used. In most cases, privacy protection is uniformly applied to all binary features resulting in data degradation and corresponding loss of performance. To avoid this undesirable effect we propose a new privacy amplification technique that is based on data hiding principles and benefits from side information about bit reliability a.k.a. *soft fingerprinting*. In this paper, we investigate the identification-rate vs privacy-leak trade-off. The analysis is performed for the case of a perfect match between side information shared between the encoder and decoder as well as for the case of partial side information.

1. INTRODUCTION

Content identification systems are widely used in various emerging applications ranging from identification of physical objects and humans to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on binary digital fingerprinting. A digital fingerprint represents a short, robust and distinctive content description allowing fast and privacy-preserving operations. In this case, all operations are performed on the fingerprint instead of on the original large and privacy-sensitive data, thus allowing introduction of crypto-based security into the analog or noisy digital world.[1] These new techniques are able to overcome the fundamental sensitivity of classical cryptographic encryption and one-way functions to minor noise in input data by trading-off the security and robustness.

This paper is an extension of our previous work.[2,3] We have previously considered the rate-privacy-complexity trade-off for identification applications.[2] This approach is based on global privacy amplification, where all bits of stored fingerprints are randomized with identical probability disregarding their reliabilities. This approach is similar in spirit to a compression-based approach.[4] However, contrarily to the previous approach a concept of bit reliability was introduced to reduce the identification complexity based on a bounded distance decoder (BDD).[5] Obviously, such a construction does not fully benefit from the fact that the information about the reliable bits can be present at the encoder and decoder, which can be used not only for the efficient decoding but also for the enhanced privacy amplification.

On accuracy, robustness and security of bag-of-word search systems

Sviatoslav Voloshynovskiy*, Maurits Diephuis, Dimche Kostadinov, Farzad Farhadzadeh and Taras Holotyak
University of Geneva, Department of Computer Science, Stochastic Information Processing Group
7 route de Drize, CH 1227, Geneva, Switzerland

ABSTRACT

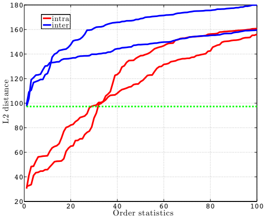
In this paper, we present a statistical framework for the analysis of the performance of Bag-of-Words (BOW) systems. The paper aims at establishing a better understanding of the impact of different elements of BOW systems such as the robustness of descriptors, accuracy of assignment, descriptor compression and pooling and finally decision making. We also study the impact of geometrical information on the BOW system performance and compare the results with different pooling strategies. The proposed framework can also be of interest for a security and privacy analysis of BOW systems. The experimental results on real images and descriptors confirm our theoretical findings.

Notation: We use capital letters to denote scalar random variables X and $\mathbf{X}$ to denote vector random variables, corresponding small letters x and $\mathbf{x}$ to denote the realisations of scalar and vector random variables, respectively. We use $\mathbf{X} \sim p_{\mathbf{X}}(\mathbf{x})$ or simply $\mathbf{X} \sim p(\mathbf{x})$ to indicate that a random variable $\mathbf{X}$ is distributed according to $p_{\mathbf{X}}(\mathbf{x})$. $\mathcal{N}(\mu, \sigma_X^2)$ stands for the Gaussian distribution with mean μ and variance σ_X^2. $\mathcal{B}(L, P_b)$ denotes the binomial distribution with sequence length L and probability of success P_b. $||.||$ denotes the Euclidean vector norm and $Q(.)$ stands for the Q-function. $\mathcal{D}(.||.)$ denotes the divergence and $\mathbb{E}\{.\}$ denotes the expectation.

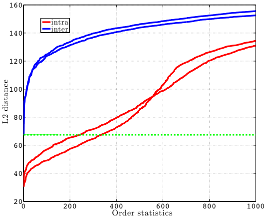
1. INTRODUCTION

The BOW framework has been widely used in content search systems, biometric applications such as face or gait recognition and more recently in multimedia security applications including copy detection, black list tracking, content blocking and commercial content ranking systems. Modern BOW based systems can easily handle large-scale search or recognition problems, even on mobile phones. The BOW approach is based on the construction of a *visual alphabet* or *dictionary* based on the clustering of low-level features such as discriminative and robust descriptors.

100 SIFTs enrolled



1000 SIFTs enrolled

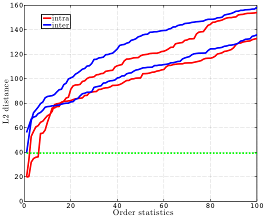


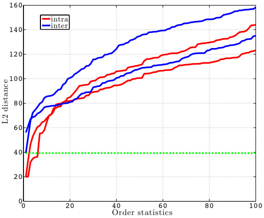
Exisiting technologies and their restrictions (5)

Example of SIFT: microstructure images (BoF without geometric re-ranking)

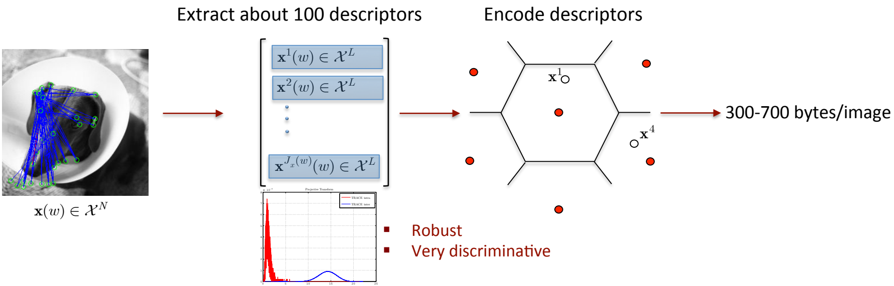
100 SIFTs enrolled

1000 SIFTs enrolled





Proposed solution: SketchPrint



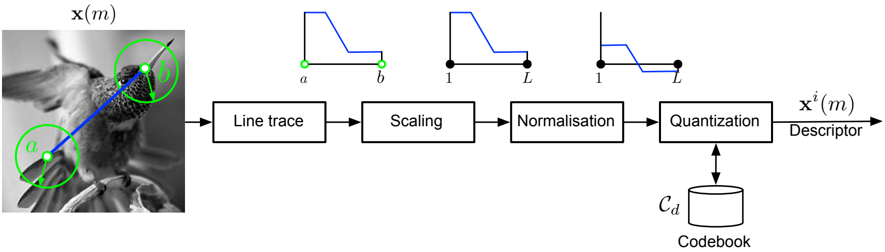
Strategy:

- To use a small number of very discriminative and robust descriptors
- No need in complex encoding (fine VQ suffices $\Rightarrow$ high precision)
- Do not store any geometric information $\Rightarrow$ memory, complexity, no need in geometric re-ranking

Sketch descriptor

SketchPrint main idea

Extract a sketch connecting two reference points



Main steps of SketchPrint:

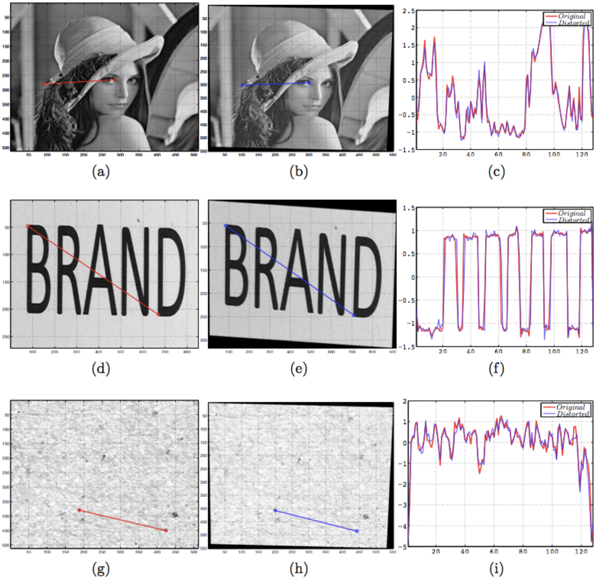
- key-points detection and filtering
- SketchPrints extraction and filtering

SketchPrint on different contents: discriminative power

Images

Text/Logos

Random microstructures



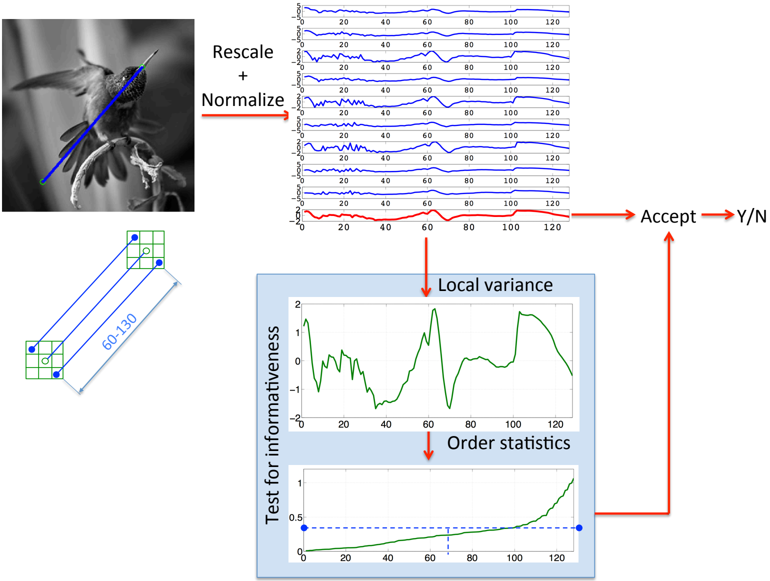
Robust key point extraction and filtering

Main problem No reliable key-point detector exists and no measure of reliability

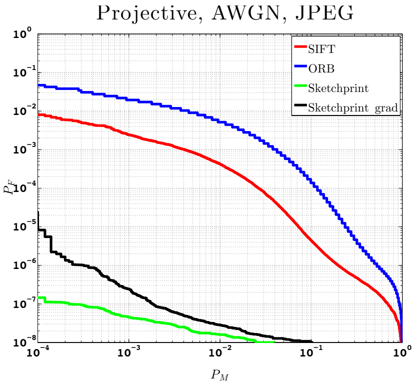
Core idea

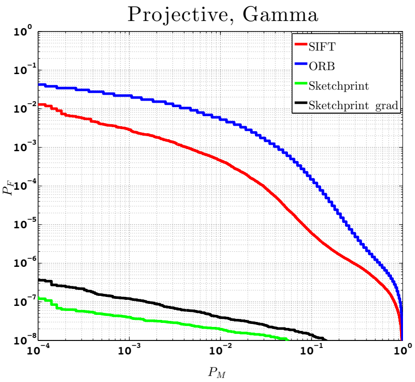
- FAST key point detector tends to produced clustered key-points under certain parameters
- Use redundancy to estimate reliability ⇒ **clustering**

SketchPrints extraction and filtering



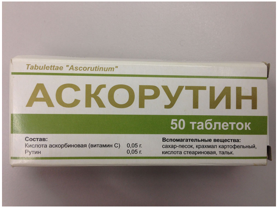
Descriptor testing: known key point positions





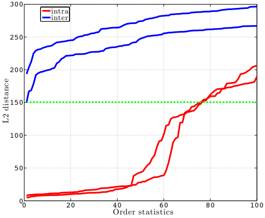
Descriptor testing: SIFT vs SketchPrint

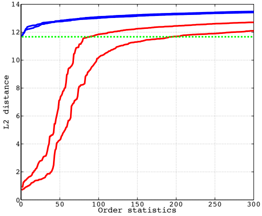
Example of SIFT: real packages (BoF without geometric re-ranking)



100 SIFTs enrolled

100 SketchPrints enrolled





Example of SIFT: text documents (BoF without geometric re-ranking)

Private content identification based on soft fingerprinting

Sviatoslav Voloshynovskiy*, Taras Holotyak, Oleksiy Koval, Fokko Beekhof and Farzad Farhadzadeh
University of Geneva, Department of Computer Science,
7 route de Drize, CH 1227, Geneva, Switzerland

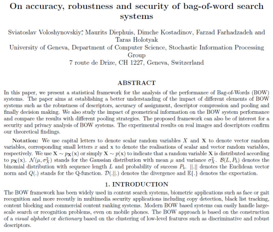
ABSTRACT

In many problems such as biometrics, multimedia search, retrieval, recommendation systems requiring privacy-preserving similarity computations and identification, some binary features are stored in the public domain or outsourced to third parties that might raise certain privacy concerns about the original data. To avoid this privacy leak, privacy protection is used. In most cases, privacy protection is uniformly applied to all binary features resulting in data degradation and corresponding loss of performance. To avoid this undesirable effect we propose a new privacy amplification technique that is based on data hiding principles and benefits from side information about bit reliability a.k.a. *soft fingerprinting*. In this paper, we investigate the identification-rate vs privacy-leak trade-off. The analysis is performed for the case of a perfect match between side information shared between the encoder and decoder as well as for the case of partial side information.

1. INTRODUCTION

Content identification systems are widely used in various emerging applications ranging from identification of physical objects and humans to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on binary digital fingerprinting. A digital fingerprint represents a short, robust and distinctive content description allowing fast and privacy-preserving operations. In this case, all operations are performed on the fingerprint instead of on the original large and privacy-sensitive data, thus allowing introduction of crypto-based security into the analog or noisy digital world.[1] These new techniques are able to overcome the fundamental sensitivity of classical cryptographic encryption and one-way functions to minor noise in input data by trading-off the security and robustness.

This paper is an extension of our previous work.[2,3] We have previously considered the rate-privacy-complexity trade-off for identification applications.[2] This approach is based on global privacy amplification, where all bits of stored fingerprints are randomized with identical probability disregarding their reliabilities. This approach is similar in spirit to a compression-based approach.[4] However, contrarily to the previous approach a concept of bit reliability was introduced to reduce the identification complexity based on a bounded distance decoder (BDD).[5] Obviously, such a construction does not fully benefit from the fact that the information about the reliable bits can be present at the encoder and decoder, which can be used not only for the efficient decoding but also for the enhanced privacy amplification.



On accuracy, robustness and security of bag-of-word search systems

Sviatoslav Voloshynovskiy*, Maurits Diephuis, Dimche Kostadinov, Farzad Farhadzadeh and Taras Holotyak
University of Geneva, Department of Computer Science, Stochastic Information Processing Group
7 route de Drize, CH 1227, Geneva, Switzerland

ABSTRACT

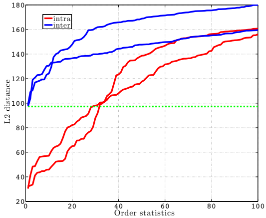
In this paper, we present a statistical framework for the analysis of the performance of Bag-of-Words (BOW) systems. The paper aims at establishing a better understanding of the impact of different elements of BOW systems such as the robustness of descriptors, accuracy of assignment, descriptor compression and pooling and finally decision making. We also study the impact of geometrical information on the BOW system performance and compare the results with different pooling strategies. The proposed framework can also be of interest for a security and privacy analysis of BOW systems. The experimental results on real images and descriptors confirm our theoretical findings.

Notation: We use capital letters to denote scalar random variables X and $\mathbf{X}$ to denote vector random variables, corresponding small letters x and $\mathbf{x}$ to denote the realisations of scalar and vector random variables, respectively. We use $\mathbf{X} \sim p_{\mathbf{X}}(\mathbf{x})$ or simply $\mathbf{X} \sim p(\mathbf{x})$ to indicate that a random variable $\mathbf{X}$ is distributed according to $p_{\mathbf{X}}(\mathbf{x})$. $\mathcal{N}(\mu, \sigma_X^2)$ stands for the Gaussian distribution with mean μ and variance σ_X^2. $\mathcal{B}(L, P_b)$ denotes the binomial distribution with sequence length L and probability of success P_b. $||.||$ denotes the Euclidean vector norm and $Q(.)$ stands for the Q-function. $\mathcal{D}(.||.)$ denotes the divergence and $\mathbb{E}\{.\}$ denotes the expectation.

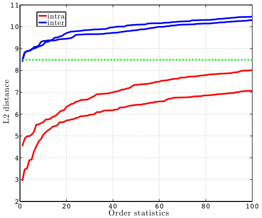
1. INTRODUCTION

The BOW framework has been widely used in content search systems, biometric applications such as face or gait recognition and more recently in multimedia security applications including copy detection, black list tracking, content blocking and commercial content ranking systems. Modern BOW based systems can easily handle large-scale search or recognition problems, even on mobile phones. The BOW approach is based on the construction of a *visual alphabet* or *dictionary* based on the clustering of low-level features such as discriminative and robust descriptors.

100 SIFTs enrolled



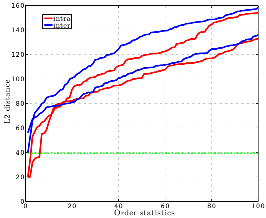
100 SketchPrints enrolled

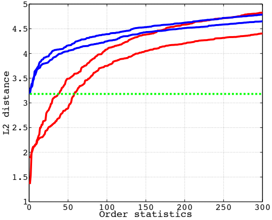


Example of SIFT: microstructure images (BoF without geometric re-ranking)

100 SIFTs enrolled

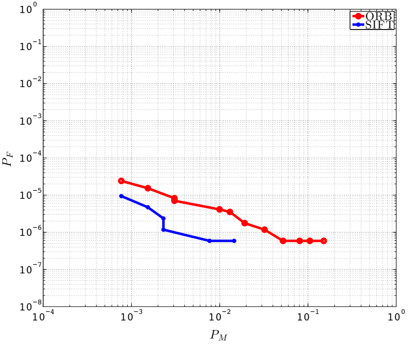
100 SketchPrints enrolled





Descriptor testing: Identification on UCID dataset

Identification test on UCID dataset: SIFT, ORB and SketchPrint real images under projective transform, AWGN ($\sigma = 10$) and JPEG Q=80



Remark:

- SketchPrint produces unique identification without any geometric re-ranking

Brand security based on "high-res" visual inspection

- Buy from eBay and enjoy your ... fake
- Can you find the differences (without the original)?
- http://www.dino.co.uk/labs/2011/how-to-spot-fake-chanel-coco-mademoiselle/

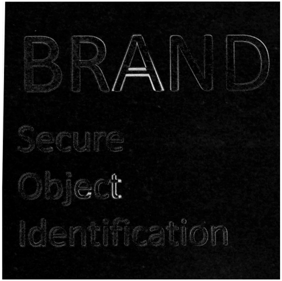
Counterfeiting detection

Once object is identified $\Rightarrow$ his design is known

Original

Fake

Detected difference



Conclusions

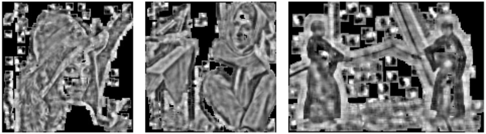
New framework

- SketchPrint works well on different visual contents
- SketchPrint is more robust, distinctive and compact than SIFT
- Efficient search and storage without any geometric re-ranking
- Potential gains for security and privacy

Counterfeiting: reconstruction from descriptors

Security leaks: the counterfeiter can learn secret features from descriptors

[P. Weinzaepfel et al, Reconstructing an Image from Its Local Descriptors, CVPR11]



[E. d'Angelo et al, From Bits to Images: Inversion of Local Descriptors, ICPR12]

SketchPrint: one can reconstruct from 1000 SIFTS with geometry but it is difficult to reconstruct from 100 SketchPrints without geometry!

The End