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

S. Voloshynovskiy, M. Diephuis, T. Holotyak

University of Geneva
Switzerland

February 11, 2015
1 Problem under consideration

2 Existing 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
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
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 \( \Rightarrow \) **huge redundancy**
- Very **complex encoding** methods are used to compress this redundancy
- **Geometric re-ranking** is needed for fine pruning
Existing technologies: SIFT+BOF+RANSAC (2)

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

S. Voloshynovskiy, M. Diephuis, T. Holotyak
University of Geneva
February 11, 2015
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)
Exisiting technologies and their restrictions (4)

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

Private content identification based on soft fingerprinting
Sviatoslav Voloshynovskiy, Tame Hekstyek, Oleksiy Koval, Folkert Beckhoff and Farhad Farhadihak
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 usually 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-enhancement 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.

2. INTRODUCTION
Content identification systems are widely used in various emerging applications ranging from identification of physical objects and human to multimedia management (content filtering, content tagging) and security (copyright protection, fraudulent monitoring, etc.). Most identification techniques are based on binary 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 the original large and privacy-sensitive data, thus allowing intrusion of crypto-based security into the analog or noisy digital world. Their new techniques are able to overcome the fundamental sensitivity of classical cryptographical encryption and one-way functions to induce noise in input data for masking the security and robustness.

This paper is an extension of our previous work. We have previously examined the no-privacy-complexity tradeoff for identification applications. This approach is based on global privacy amplification, where all bits of stored fingerprints are modulated with identical probability disengaging their reliabilities. This approach does not rely on a compression-based approach. However, normally to the previous approach a concept of bit reliability was introduced to reduce the identification complexity based on a bounded distance decoder (BDD). 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, Dinhac Kostadinov, Farhad Farhadihak and Tame Hekstyek
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 choices of BoW systems such as the robustness of descriptors, accuracy of assignment, descriptor aggregation and pooling and finally the impact of the external information on the BoW system performance and compare the results with different weighting strategies. The proposed framework can also be of interest for security and privacy analysis of BoW systems. The experimental results on real images and descriptors conform to theoretical findings.

Notations. We use capital letters to denote scalar random variables X and x to denote vector random variables, corresponding small letters x and X to denote the realizations of scalar and vector random variables, respectively. We use N(μ, σ²) or simply N(μ, σ²) to indicate that a random variable X is distributed according to N(μ, σ²). N(μ, σ²) stands for the Gaussian distribution with mean μ and variance σ². \( d_2 \) denotes the Euclidean distance of two vectors and \( L_2 \) stands for the Q-function. \( \tanh \) denotes the hyperbolic tangent and \( \tanh \) denotes the exponential.

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 such as detection, block for block, control blocking and commercial content matching systems. Modern BoW systems can easily handle large-scale search and retrieval 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.
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 ⇒ high precision)
- Do not store any geometric information ⇒ memory, complexity, no need in geometric re-ranking

Extract about 100 descriptors

\[ x(w) \in \mathcal{X}^N \]

\[ x^1(w) \in \mathcal{X}^L \]
\[ x^2(w) \in \mathcal{X}^L \]
\[ \vdots \]
\[ x^J(w) \in \mathcal{X}^L \]

Encode descriptors

\[ 300-700 \text{ bytes/image} \]

- Robust
- Very discriminative
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 produce clustered key-points under certain parameters
- Use redundancy to estimate reliability ⇒ clustering
SketchPrints extraction and filtering

Rescale + Normalize

Local variance

Order statistics

Test for informativeness

Accept Y/N

S. Voloshynovskiy, M. Diephuis, T. Holotyak
Descriptor testing: known key point positions

Projective, AWGN, JPEG

Projective, Gamma

S. Voloshynovskiy, M. Diephuis, T. Holotyak
Descriptor testing: SIFT vs SketchPrint

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

100 SIFTs enrolled  100 SketchPrints enrolled
Descriptor testing: SIFT vs SketchPrint

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

Private content identification based on soft fingerprinting

Svitoslav Voloshynovskiy, Tane Holotyak, Oleksiy Kosen, Folklo Beckhof and Farnad Farshadzadeh
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 usually 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 a data hiding primitive 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.

2. INTRODUCTION

Content identification systems are widely used in various emerging applications ranging from identification of physical objects and biometric identification to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on binary fingerprinting. A digital fingerprint represents a short, robust and distinctive content descriptor allowing fast and privacy-preserving operation. 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 many digital worlds. These new techniques are able to overcome the fundamental sensitivity of classical cryptographic encryption and one-way functions to generic noise in input data thus breaking the security robustness.

This paper is an extension of our previous work. We have previously analyzed the non-privacy-complexity tradeoff for identification applications. This approach is based on global privacy amplification, where all bits of stored fingerprints are modified with identical probability diminishing their reliabilities. This approach is similar in spirit to a compression-based approach. However, contrary to the previous approach a concept of bit reliability was introduced to reduce the identification complexity based on a bounded distance decoder (BDD). Obviously, each 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.

100 SIFTs enrolled

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

Svitoslav Voloshynovskiy, Maurice Diephuis, Doina Kostadinova, Farnad Farshadzadeh and Tane Holotyak
University of Geneva, Department of Computer Sciences, 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 choices of BOW systems such as the robustness of descriptors, accuracy of assignments, descriptor aggregation and pooling and finally distance metrics. We also study the impact of precomputed 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.

Notations. We use capital letters to denote scalar random variables X and x to denote vector random variables, corresponding small letters x and x to denote the realizations of scalar and vector random variables, respectively. We use N − p, N(p) or (N, p) to indicate that a random variable X is distributed according to p(x). N(x, d) stands for the Gaussian distribution with mean x and variance d. B(x, L) stands for the binomial distribution with sequence length L and probability of success p. (.) stands for the Euclidean vector norm and |.| denotes the absolute value. 

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 such as covert communication, data-hiding, content blocking and commercial content rating systems. Modern BOW-based systems can easily handle large-scale search or content protection tasks even on mobile devices. 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.
Descriptor testing: SIFT vs SketchPrint

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
Buy from eBay and enjoy your ... fake
Can you find the differences (without the original)?
Once object is identified $\Rightarrow$ his design is known
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

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