SKETCHPRINT: PHYSICAL OBJECT MICRO-STRUCTURE IDENTIFICATION USING MOBILE PHONES

Maurits Diephuis, Sviatoslav Voloshynovskiy, Taras Holotyak

Stochastic Information Processing Group, University of Geneva
7, Route de Drize, CH-1227, Carouge, Switzerland

ABSTRACT

This paper addresses the identification of physical objects based on their physical non-cloneable surface structures. These micro-structures are optically acquired using a hand held non-modified consumer mobile phone.

Object identification is done with the *SketchPrint* descriptor, which combines fingerprint-like properties while having reasonable invariance to geometrical and lighting distortions due to its semi-local nature. Crucially, objects can be identified without any geometrical matching or final re-ranking procedure.

Index Terms— physical object identification, micro-structure images, semi-local descriptor, SketchPrint

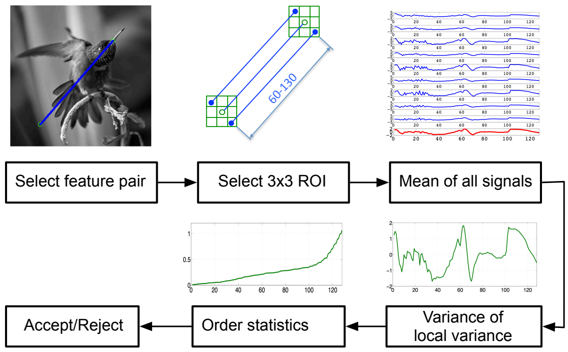

Fig. 1: Basic work-flow of the SketchPrint descriptor algorithm.

1. INTRODUCTION

Identification of physical objects based on physical non cloneable functions (PUF's) such as optically acquired micro-structures images is an attractive solution for many domains, e.g. anti-couterfeiting or tracking and tracing. [1] Similar to biometrics, the working principle is the non-cloneable character of the micro-structures. Architectures based on micro-structures are cheap to enrol, do not require any product modification and verification can be done relatively easy.

Key elements in the processing chain are the selection of the relevant image patch from which the micro-structure is extracted, the selection of robust features and finally different compression techniques resulting in a short fingerprint.

The main challenge consists in the design of proper descriptors for micro-structure images and the corresponding identification framework. Classical celebrated local descriptors such as SIFT [2] or ORB [3] which were developed for natural images are characterized by very distinctive features of image patches in the vicinity of key points such as corners. Applied to micro-structure images these local descriptors fail to produce a robust and discriminative image representation due to the "uniformity" of micro-structure image statistics. This explains their low discriminative power. Moreover, the lack of distinctive elements leads to weak stability of detected key-point and poor geometric consistency. Finally, the number of descriptors generated per typical micro-structure image is about 10'000-40'000 thus potentially exceeding the size of the original image. Obviously, descriptors which should identify a large amount of images need to be as compact as possible.

In this paper, we try to overcome the above problems by proposing a new type of image representation in the form of SketchPrint (Figure 1) [4]. Having demonstrated very good robustness and discrimination properties for natural images and text documents [4], we intend to extend SketchPrint to the identification of micro-structure images acquired by mobile phones. The key idea behind SketchPrint is to form a more discriminative descriptor, that can function as a persistent hash, of which about 100-200 need to be enrolled per image. Moreover, such kind of representation is quite unique and does not require any additional geometric re-ranking thus providing a memory efficient content representation and fast identification.

This paper is organized as follows: we consider the related work in Section 2, Section 3 presents the SketchPrint descriptor and identification architecture. The experimental results are presented in Section 4 and Section 5 concludes the paper.

Notation: We use capital letters to denote scalar random variables X and $\mathbf{X} = \{X[1], X[2], \ldots, X[N]\}$ to denote vector random variables, corresponding small letters x and $\mathbf{x} = \{x[1], x[2], \ldots, x[N]\}$ to denote the realisations of scalar and vector random variables, respectively. All vectors are assumed to be of the length N. $||.||$ denotes Euclidean vector norm. A descriptor with the index k from an image with index w is denoted by $\mathbf{x}^k(w)$ and its individual elements are denoted as $x_i^k(w)$ with $1 \leq i \leq N$.

2. RELATED WORK AND REQUIREMENTS

Previous work on micro-structures, predominantly on the public FAMOS dataset [1] [1] focused on the following methodologies. In [1] a known printed template was used as a mark to guide the extraction of the correct image patch and to compensate any geometrical distortions. Results in [5] show that micro-structure images may also be authenticated using traditional computer vision features at a computational cost as images are exhaustively matched based on their feature geometry and an outlier detection algorithm. Although, this approach demonstrates the feasibility of reliable identification, it is not practically attractive due to the huge complexity.

Therefore, we will focus our analysis on an approach that produces the state-of-the-art results in general applications [6] and is based on the Bag-of-Features (BoF) and RANSAC based geometrical re-ranking. A block diagram of this architecture is shown in Figure 2. A probe image $\mathbf{y}$ is presented to the identification system which extracts local descriptors like SIFT. To reduce the complexity of RANSAC based geometric matching of local features between the probe and enrolled templates, the BoF module of identification systems produces a short list $\mathcal{L}(\mathbf{y})$ of the most likely candidates whose encoded local features match the best to the probe ones. The encoding includes a proper aggregation of features to produce a short and memory efficient yet discriminative image representation. The geometric re-ranking module works only with the list of templates $\mathcal{L}(\mathbf{y})$ and select such a template index $\hat{w}$ which attains the largest number of geometrically consistent matching according to the assumed model of geometric transformations typical for the mobile phone acquisitions. Although widely used, this approach faces issues when applied to the identification of micro-structure images: 1) The number of descriptors should be very high to produce a good consistency between the descriptors of probe and enrolled items. Otherwise, RANSAC fails to reliably match them. For a small amount of points selected according to a certain description pre-selection rule, no geometric consistence is observed due to an unacceptable amount of outliers which RANSAC cannot handle. 2) The large amount of descriptors (10'000-40'000) and their geometrical positions require significant storage , thus an efficient compression and aggregation are needed. 3) The compression/aggregation are information lossy operations that reduce the distinguishability of BoF. As a result the efficiency of BoF is drastically reduced since almost all images look "similar" and the only possibility to distinguish them is with geometric matching.

[1] `http://sip.unige.ch/famos`

This paper continues on the work in [4] where SketchPrint was conceptually introduced and used in a Bag-of-Features framework without quantization and just storing 100 SketchPrints per image without any further re-ranking. This article will apply SketchPrint to a realistic set of micro-structures with a new fusion and decision framework based on ordered statistics.

3. SKETCHPRINT BASED IDENTIFICATION

The core of the identification is the SketchPrint descriptor, computation of which for a pair of key-points, is schematically shown in Figure 1. It consists of two main stages: the detection and filtering of key-points and extraction and filtering of descriptors.

3.1. Key-point detection and filtering

As SketchPrint traces are formed between a pair of two key-points, the stability of those points is much more crucial than for algorithms where the descriptor is determined around the region of a single feature point. Therefore, it is vital that the probability of miss should stay as small as possible. The algorithm detects key-points as follows:

- Micro-structure images are enhanced using local histogram equalisation.
- The ORB key-point detector is used to detect an approximate set of points [3] .
- Detected key-points are spatially clustered using graph connected components. Clusters of local points are spatially averaged to form a new single points, while clusters with a single point are removed. This step discards the scale-space a key-point was detected in.

3.2. Descriptor extraction and filtering

The SketchPrint descriptor is obtains as follows:

- Feature points are chosen as pair when the distance between them is between the number of used interpolation points, i.e., between L_d and $2L_d$, where $L_d = 256$.
- For each point of a pair, a 3×3 neighbourhood is taken, after which the image is interpolated 9 times.
- Signals are re-scaled to a fixed length, i.e., 128 or 256.
- All these signals are averaged to a single SketchPrint.
- SketchPrints are re-normalised to produce a zero-mean unit variance vector.
- The most informative SketchPrints are chosen by only selecting those whose variance of local variances are at least 1.5 times higher than the average for that specific image.
- SketchPrints whose variance of local variances is disproportionally caused by running over an edge giving a single

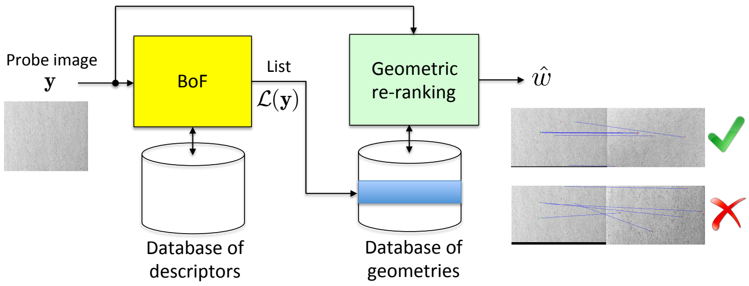

Fig. 2: Typical BoF+RANSAC image identification architecture.

large peak, or step, over an otherwise flat signal, are also rejected using ordered statistics.

- The final set of SketchPrints is written out, without any other (geometrical) information.

3.3. Identification architecture

A database $\mathcal{D}$ of enrolled feature vectors contains a collection of M items, represented by their descriptors $\{\mathbf{x}^k(w)\}$, $1 \le w \le M$ and $1 \le k \le J_x(w)$, where $J_x(w)$ denotes the number of descriptors for an item w.

The identification problem is to find the best match between a query, represented by a set of descriptors $\{\mathbf{y}^j\}$, $1 \le j \le J_y$, and those in the database $\mathcal{D}$. The system should produce an index estimate $\hat{w}$ for the best match, or an empty set $\emptyset$ if the query is not related to the database. Besides the possibility to produce a list of best matches, in this paper we only consider unique decoding.

The proposed architecture can be seen in Figure 3. A probe descriptor $\{\mathbf{y}^j\}$, $1 \le j \le J_y$, is matched against the database $\mathcal{D}$ by computing the distance $d_{kj}(w) = ||\mathbf{x}^k(w) - \mathbf{y}^j||^2$. It is important to point out that any fast approximate distance computation, like those based on product vector quantization [6] and advanced aggregation techniques such as Fisher vectors [7] or difference vectors [8] may be used. However, our goal is to demonstrate the feasibility of identifying random micro-structures based solely on descriptors without any further geometric processing and re-ranking.

Given the distances $d_{kj}(w)$, one can consider several system designs such that for each descriptor $\mathbf{y}^j$ the system returns: (a) the whole set of distances for all descriptors stored in the database $\mathcal{D}$; (b) only the list of descriptors and their indices that are within some ϵN from $\mathbf{y}^j$, i.e., ϵ-NN list $\mathcal{L}(\mathbf{y}^j) = \{w : d_{kj}(w) \le \epsilon N\}$, or the closest ℓ-NN descriptors and their indices; or (c) the maximum likelihood (the closet match) when $\ell = 1$ or equivalently $\hat{w} = \text{argmin}_{w,k}\, d_{kj}(w)$; To investigate the theoretical performance limits we will consider the case (a), which is characterized by the minimum probability of miss, leaving aside the memory/complexity issues and assuming that a reasonable low-complexity approximation can be obtained with case (b) and in some set-ups with case (c). This depends on the descriptor robustness and distinguishability.

Independent of the approach used, all distances computed above for all J_y descriptors are combined to a common stack and the $L \le J_y$-smallest distances $\tilde{d}_r(w)$, $1 \le r \le L$, such that $\tilde{d}_1(w) \le \tilde{d}_2(w) \le \cdots \le \tilde{d}_L(w)$, are produced based on the analysis of distance order statistics (DOS) as shown in Figure 3.

The identification decision rule is based on the order statistic detector, which can produce both a soft and hard decision. The latter is defined as:

$$D_r(w) = \begin{cases} 1, & \text{if } \tilde{d}_r(w) \ge T_r, \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

using a set of trained thresholds $\{T_r\}$, $1 \le r \le L$ for each order statistic r. These thresholds are set specifically for our application domain, such that the probability of miss for the rth DOS is bounded as $P_{M_r} \le \epsilon'$, with ϵ' to be a small non negative constant, since SketchPrints are discriminative in nature and thus naturally exhibit small values of probability of false acceptance P_{FA_r}. It is important that descriptors are not missed at the decision stage. The decision outputs $D_r(w,k)$ may also be soft values computed proportionally to the statistics of correct and incorrect DOSs. This option will be considered in future research.

Following the BoF+RANSAC strategy (Figure 2) the framework should produce a list $\mathcal{L}(\mathbf{y})$ of the most likely candidates which can then be geometrically re-ranked. The corresponding list decoder is:

$$\mathcal{L}(\mathbf{y}) = \{w : s(w) \ge T\}, \quad (2)$$

where

$$s(w) = \sum_{r=1}^{L} D_r(w). \quad (3)$$

Note that for simplicity this models the individual order statistics as independent whereas in reality they are dependent. System performance may then be evaluated by the probability of missing a correct item w:

$$P_M = \Pr[S(w) \le T \mid w], \quad (4)$$

for a chosen threshold T and the probability of falsely accepting a non-related item as matching to an enrolled item in the database:

$$P_{FA} = \Pr[S(w) > T \mid w'] \quad w' \neq w. \tag{5}$$

The resulting average list size of retrieved indices is then:

$$\mathbb{E}\{|\mathcal{L}(\mathbf{y})|\} \simeq MP_{FA}, \tag{6}$$

assuming the probability of miss is selected as $P_M \leq \epsilon$.

4. EXPERIMENTAL VALIDATION

For the empirical test, the micro-structure images of 50 paper sheets were photographed with a hand held Samsung Galaxy III in ordinary light, 5 times each, roughly in the same positions and without any geometric templates. An example of the resulting (cropped) images can be seen in Figures 4a-4d.

Every image was enrolled once, (a) with a 100 SketchPrint descriptors; (b) with a 1'000 SIFT via gradient magnitude and (c) an unlimited regime with on average 10'000 SIFT descriptors per image [9].

This enrolled set was tested against the other acquired images, that functioned as query. At query side, the number of descriptors was not constrained resulting in on average 2'000 SketchPrint descriptors and 10'000 SIFT descriptors per query image.

A working example is shown in Figure 4. The intra class is formed from the ordered statistic distances from descriptors originating from identical images and the inter class distances for those between non-identical images.

The results for the whole dataset are shown in Figure 5. It should be clear, as the intra and inter first order distances for SketchPrint do not overlap, that for this set, error-less identification is possible even without additional decoding rules.

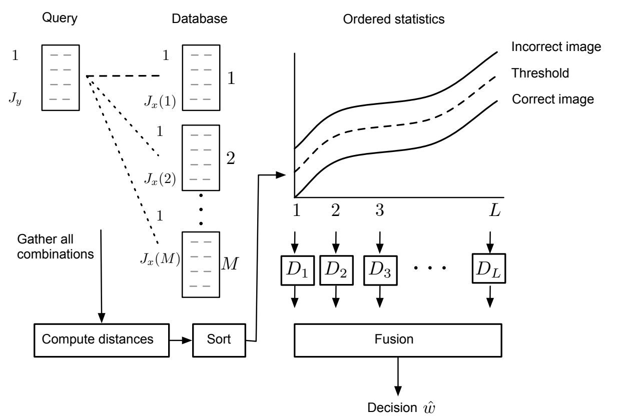

Fig. 3: Identification architecture based on SketchPrint and ordered statistics.

Table 1 shows the average retrieved list size $|\mathcal{L}|$ when the decoding rule (3) uses 1 to 10 cumulative order statistics and $P_M = 0$.

Error-less identification ($|\mathcal{L}| = 1$) is not possible for SIFT in this particular framework, which can be explained by the fact that its feature-points are not stable enough, nor its descriptors discriminative enough for micro-structure images.

5. CONCLUSION

In this work, we have detailed the SketchPrint algorithm and shown the feasibility of using this framework for identifying objects via their micro-structure using a mobile phone.

Future work will deploy quantization techniques on top of the original SketchPrints and use and test different decision frameworks using a larger amount of images.

Acknowledgement: This paper was partially supported by SNF project 200020-146379. Corresponding author is S. Voloshynovskiy, `svolos@unige.ch`.

REFERENCES

[1] S. Voloshynovskiy, M. Diephuis, F. Beekhof, O. Koval, and B. Keel, "Towards reproducible results in authentication based on physical non-cloneable functions: The forensic authentication microstructure optical set (famos)," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5 2012.

[2] M. Brown and D. Lowe, "Invariant features from interest point groups," 2002.

[3] E. Rublee, Vincent V. Rabaud, K. Konolige, and G. Bradski, "Orb: an efficient alternative to sift or surf," in *Computer Vision (ICCV), 2011 IEEE International Conference*. IEEE, 2011, pp. 2564–2571.

[4] S. Voloshynovskiy, M. Diephuis, and T. Holotyak, "Mobile visual object identification: from sift-bof-ransac to sketchprint," in *Proceedings of SPIE Photonics West, Electronic Imaging, Media Forensics and Security V*, San Francisco, USA, January, 13 2015.

[5] M. Diephuis, S. Voloshynovskiy, T. Holotyak, Nabil Standardo, and Bruno Keel, "A framework for fast and secure packaging identification on mobile phones," in *Proceedings of SPIE Photonics West, Electronic Imaging, Media Forensics and Security V*, San Francisco, USA, January, 23 2014.

[6] J. Hervé, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference*. IEEE, 2010, pp. 3304–3311.

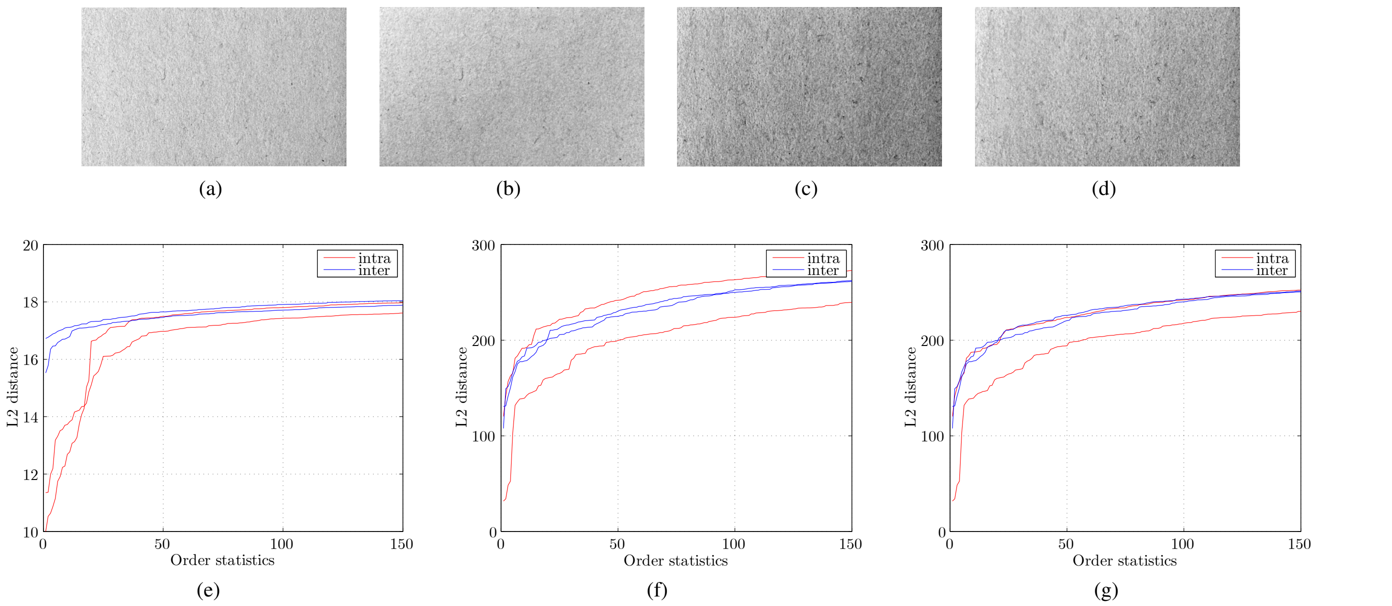

Fig. 4: Order-statistic example for a two acquisitions of two micro-structures using SketchPrint and SIFT: (a-b) a sample 1 and (c-d) a sample 2 both acquired 2 times, (e) the DOS for 100 enrolled SketchPrint, (f) the DOS for 1’000 SIFT and (g) 10’000. SIFT. The latter can not distinguish these micro-structures due to the overlap of inter and intra statistics.

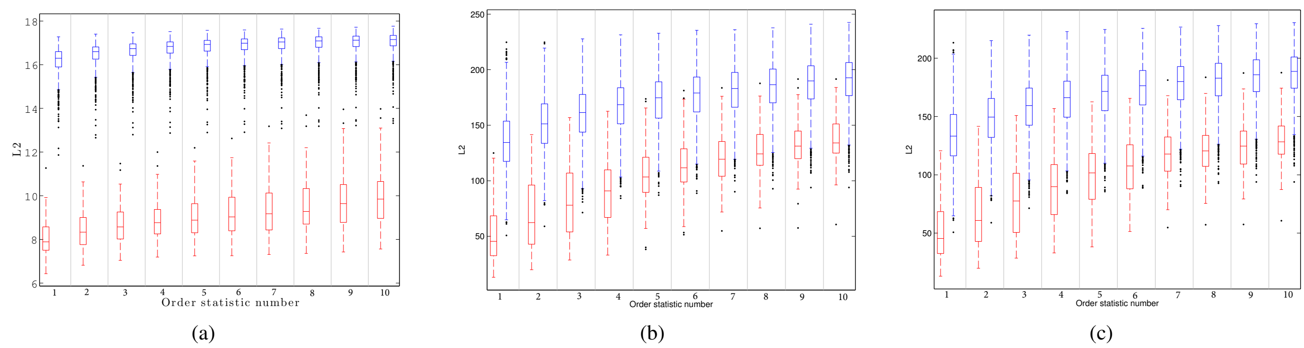

Fig. 5: Boxplot for the first 10 distance order statistics for intra (red) and inter (blue) distances: (a) SketchPrint, (b) 1’000 enrolled SIFT and (c) 10’000 enrolled SIFT based identification.

	Average retrieved list size $\lvert\mathcal{L}\rvert$									
Order Statistic	1	$1 \div 2$	$1 \div 3$	$1 \div 4$	$1 \div 5$	$1 \div 6$	$1 \div 7$	$1 \div 8$	$1 \div 9$	$1 \div 10$
SketchPrint	1	1	1	1	1	1	1	1	1	1
SIFT 1’000	18	14	13	12	12	12	12	12	12	11
SIFT 10’000	16	13	12	11	11	10	10	10	10	10

Table 1: The average retrieved list size $\lvert\mathcal{L}\rvert$ per cumulatively used order statistic (3), for SketchPrint and SIFT.

[7] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier, “Large-scale image retrieval with compressed fisher vectors,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference*. IEEE, 2010, pp. 3384–3391.

[8] B. Girod, V. Chandrasekhar, D. M. Chen, N. Cheung, R. Grzeszczuk, Y Reznik, G. Takacs, S. Tsai, and R. Vedantham, “Mobile visual search,” *Signal Processing Magazine, IEEE*, vol. 28, no. 4, pp. 61–76, 2011.

[9] A. Vedaldi and B. Fulkerson, “Vlfeat: An open and portable library of computer vision algorithms,” 2008.