Fine-grained recognition of physical objects on mobile phones: from categorization to identification

Maurits Diephuis
Stochastic Information Processing
University of Geneva, Switzerland
maurits.diephuis@unige.ch

Sviatoslav Voloshynovskiy
svolos@unige.ch

Taras Holotyak
taras.holotyak@unige.ch

Abstract

This paper addresses the recognition of physical objects such as pharmaceutical and cosmetic packages with two hierarchical levels. The first level categorises the object based on printed images, graphics and text, while the second one identifies it based on so called micro-structures: the physical non-cloneable surface structure images.

Both levels deploy a tracing 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.

1. Introduction

In this paper we consider fine-grained identification of physical objects, such as packages, in two levels. The first level categorises the type of package based on visual cues such as design elements or text. The second stage identifies an individual object based on optically acquired micro-structures which are physical non cloneable functions (PUF's). Similar to biometrics, the working principle is their unique and non-cloneable character. Such architectures are cheap to enrol, do not require any product modification and verification can be done relatively easy.

In both the categorisation and the identification we use a descriptor that traces over the image between selected key-points. The key idea is to build a discriminative descriptor, that can function as a persistent hash that does not require any exhaustive geometric re-ranking.

2. Framework

The targeted class of objects is characterized by localized graphical features that can be detected for fine-grained recognition. Such a localization increases both the accuracy of recognition and robustness to geometric distortions typical for mobile imaging.

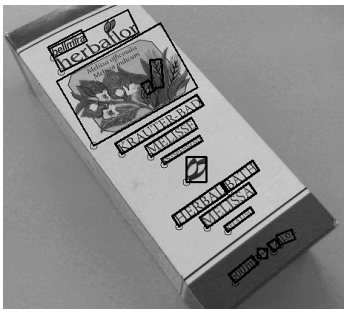

Figure 1: Automatic partitioning by the categorization algorithm. It is language and font independent.

The tracing descriptor is used twice throughout the hierarchy, but with a different method to generate the key-points to trace between. As traces are formed between a pair of two key-points, the stability of those points is crucial.

For initial categorisation of a physical package, all text and individual design elements are segmented using local variance and morphological operations to find individual elements and regions (Fig. 1). The centroids of individual graphical elements within segmented regions form the key-points that will be traced between. This enables accurate and fast categorization of an object using advanced indexing techniques or even direct fast Hamming distance based matching.

Having recognized the type of the presented object, the system proceeds with the specific identification of that object based on local micro-structures. Given the ascertained type of the object, image patches from preselected regions, free from graphical elements, are extracted.

FAST [2] key-points are detected and 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.

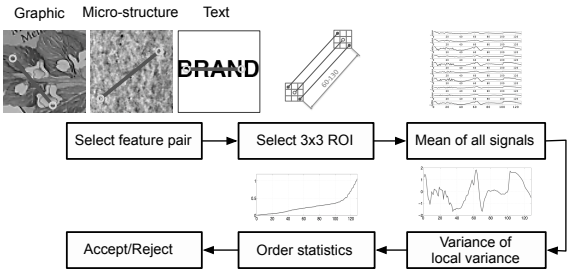

Figure 2: Basic work-flow of the tracing descriptor.

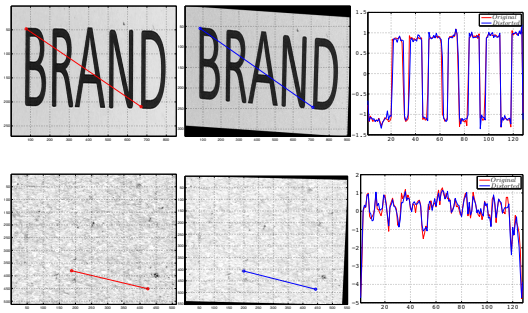

Figure 3: Trace examples from an original package and micro-structure and their distorted copies.

It stems from the observation that whole clusters of FAST key-points tend to be more stable.

Generating the trace (Fig. 2 and 3) is identical for both stages in the hierarchy and done as follows:

- Feature points are chosen as pair when the distance is sufficient given the interpolation used.
- For each point of a pair, a 3×3 neighbourhood is taken, after which the traces are computed between two sets of key points. The resulting set is rescaled and averaged in a single re-normalized signal.
- The most informative traces are chosen by only selecting signals whose variance of local variances is at least 1.5 times higher than the average for that specific image.
- Candidate traces for which this value is disproportionally caused by an edge are also rejected.
- Resulting traces are quantized using either random projections and sign or product VQ with a pre-trained codebook with about 2K centroids for each VQ block.
- Identification is based on indexing, and matches each descriptor of a probe image. The final decision is then based on majority voting.

3. Experimental validation

To empirically test the accuracy of the categorization we have constructed a database of packages and

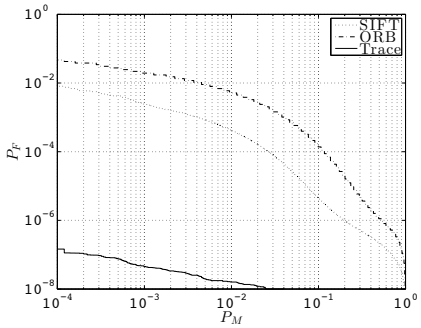

Figure 4: ROC showing the discriminative power of ORB, SIFT and Trace descriptors detected in images distorted by AWGN, JPEG compression and a projective transformation.

acquired them 3 times using a hand held Samsung Galaxy III in ordinary light. We used a single frontal image for the enrolment and the two others for the categorization testing. In total 450 packages were enrolled form which about 30 regions from each package were extracted that were passed to the trace descriptor. These were classified error-less.

To validate the identification, the micro-structure images of 50 packages were acquired with a mobile phone, 3 times each, all in about the same position of roughly 4×4 cm regions. Every image was enrolled, (a) with a 100 trace descriptors; (b) with a 1'000 SIFT descriptors selected via gradient magnitude and (c) an unlimited regime with on average 10'000 SIFT descriptors per image [1]. Even on (distorted) micro-structures the trace descriptor is discriminative enough to allow for error-less identification. This is not the case for SIFT (or ORB) in this particular framework, as its feature-points are not stable enough, nor its descriptors discriminative enough. As emphasized by the ROC in Fig. 4 the probability of false acceptance for trace is significantly lower than those for SIFT and ORB.

4. Conclusion

We have shown the viability of categorizing and identifying physical objects using a simple descriptor and a mobile phone. Specifically, this can be done without a geometry based re-ranking procedure. This work was partially supported by SNF project No. 200020-146379.

References

[1] M. Brown and D. G. Lowe. Invariant features from interest point groups. In *BMVC*, number 1, 2002.

[2] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Computer Vision–ECCV 2006*, pages 430–443. Springer, 2006.