Physical object identification based on FAMOS microstructure fingerprinting: comparison of templates versus invariant features

Maurits Diephuis, Sviatoslav Voloshynovskiy, Fokko Beekhof

University of Geneva, 7 route de Drize, CH 1227, Geneva, Switzerland

[maurits.diephuis,svolos, fokko.beekhof]@unige.ch

***Abstract*—In this paper, we address the problem of physical object identification based on optical non-cloneable surface microstructure images. Physical object identification is an emerging problem raised in mobile multimedia applications that interact with physical objects as well as in physical world security applications for which there is a great need for reliable, fast and secure object verification. One of the most crucial problems in the design of identification systems is optimal feature selection and extraction which are characterised by their high distinguishability and robustness to lightening variations and geometrical transforms. Not less an important aspect of feature selection is their vulnerability to counterfeiting or physical cloning that we refer to as physical security. Since the geometric de-synchronization represents one of the most significant challenges in the design of reliable physical object identification/authentication systems, we will investigate this problem using two techniques that are well established in computer vision applications and compare the performance of both systems. In particular, we consider two different strategies based on special graphical marks present on physical objects such as packaging or watches which can be considered as templates and microstructure features extracted based on the popular SIFT descriptors. To evaluate the performance of both approaches we use the FAMOS database which contains 5000 unique carton packages acquired 6 times each with two different cameras. The performance of the systems is evaluated based on the empirically ascertained probabilities of miss and false acceptance.**

I. Introduction

The identification of physical objects, such as passports, medicine packages or luxury items, based on microstructure images represents a challenging physical world security problem. At the core of the identification architecture lie the optical acquired microstructures, which just like human biometrics, exhibit uniqueness and a fundamentally non-cloneable character. Such a microstructure based architecture ensures cheap enrolment of physical samples, offers non-invasive protection and leads straightforward and fast verification. An example can be seen in Figure 1.

The key two elements of such a system are the selection of robust or invariant features, and the extraction of the relevant microstructure image patch from an acquired raw image.

Selected features should fulfil a number of conflicting requirements. They should uniquely identify an object whilst being as robust as possible to lighting, resolution and geometric changes.

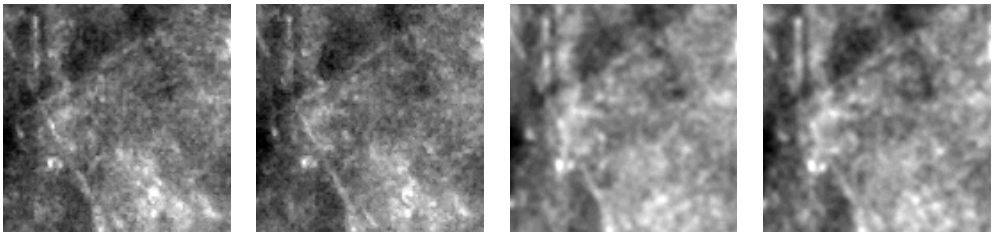
In this paper we compare two identification architectures working on real world data that has been acquired under realistic circumstances. Different camera's and lighting were used for enrolment and verification, furthermore, the acquired samples are all geometrically misaligned. In particular, we study an identification system based on a template in the form of a printed mark and a second system based on invariant (SIFT) features. The template based method synchronises all presented images against a known template using a printed mark for guidance. This ensures that samples are geometrically aligned and that a microstructure can be extracted from a predefined region. Identification is then based on the similarity between the extracted query microstructure and the synchronised microstructures in the database. Contrarily the feature based identification method does not attempt to undo the geometric distortions. Features are gathered from both the query image and the enrolled images. The system attempts to match features based on their descriptors and evaluates whether or not an affine transform can be ascertained between the query features and the dataset features. If this relation can be established, identification is deemed successful.

This paper is organized as follows. Section II provides a problem formulation and gives a system overview. Template based identification is considered in Section III, while Section IV addresses feature based identification. Both systems their performance is evaluated on the FAMOS dataset [1] in Section V. Finally, Sections VI concludes the paper.

Notation: Scalar random variables are designated by capital letters X, and bold capitals $\mathbf{X}$ denote vector random variables. Corresponding small letters x and $\mathbf{x}$ denote their respective realizations. $\mathcal{X}$ denotes the set of possible values of x. The descriptor with the index k from image with the index m is denoted by $\mathbf{x}^k(m)$, its individual elements are denoted as $x_i^k(m)$.

II. Problem formulation and system overview

Under the identification problem we assume that the enrolled database contains images acquired from M objects. We suppose that each image of size $N_1 \times N_2$ is lexicographically ordered into a sequence $\mathbf{x}(m)$ of length $N = N_1 \times N_2$ with $1 \leq m \leq M$. The object that is to be tested is represented by its own vector $\mathbf{y}$ which might originate from the observation of some $\mathbf{x}(m)$ contained in the database through the degradation channel $p(\mathbf{y}|\mathbf{x}(m))$ or any randomly generated vector $\mathbf{x}'$ that is not correlated with data stored in the database.



(a) (b) (c) (d)

Figure 1: Multiple acquisitions of a single microstructure sample from different cameras designated as Cam 1 (1a, 1b) and Cam 2 (1c, 1d). Histogram equalization was used for visualisation purposes.

The identification problem can be considered as a composite $M+1$ hypothesis testing [2], [3]:

$$\begin{cases} H_0 & : p(\mathbf{y}|H_0) = p(\mathbf{y}|\mathbf{x}'), \\ H_m & : p(\mathbf{y}|H_m) = p(\mathbf{y}|\mathbf{x}(m)), 1 \leq m \leq M, \end{cases} \quad (1)$$

where the hypothesis H_0 corresponds to the case when some to the database unrelated object $\mathbf{x}'$ is presented to the system, and hypothesis H_m denotes a valid case where the object under consideration corresponds to enrolled data item $\mathbf{x}(\mathbf{m})$, $1 \leq m \leq M$.

To validate the system's performance, we will use the *probability of successful attack* P_{sa} which denotes the case when an unrelated object is accepted as one of enrolled objects, i.e., the attacker succeeded with the counterfeiting. Secondly, we define the *probability of incorrect identification* P_{ic} which denotes the situation in which the enrolled object with the index m is wrongly decoded.

We will assume that for each enrolled object, the system is based on some optimal binary decision rule $\phi : \mathcal{Y}^N \times \mathcal{X}^N \mapsto \{0, 1\}$, where 0 stands for rejection and 1 for acceptance. In particular, the decision about the acceptance or rejection of some probe is performed according to the comparison of the corresponding distance with the threshold γ as $\phi(\mathbf{y}, \mathbf{x}(m)) \leq \gamma N$, i.e., if $\phi(\mathbf{y}, \mathbf{x}(m)) \leq \gamma N$, the decision is 1 and 0, otherwise.

In this case, the probability P_{sa} can be defined as:

$$P_{sa} = \Pr[\cup_{m=1}^{M} \phi(\mathbf{y}, \mathbf{x}(m)) \leq \gamma N | H_0]. \quad (2)$$

The probability P_{ic} consists of probabilities of two possible events under the hypothesis H_m, i.e., the probability that the object with the index m is not found in the database consisting of M objects, denoted as *probability of miss* P_m and the probability that the object with the index m is falsely accepted or identified with the index m', denoted as the *probability of false acceptance* P_f. These two probabilities are defined as:

$$P_m = \Pr[\phi(\mathbf{y}, \mathbf{x}(m)) \geq \gamma N | H_m], \quad (3)$$

$$P_f = \Pr[\cup_{m \neq m'} \phi(\mathbf{y}, \mathbf{x}(m)) \leq \gamma N | H_m], \quad (4)$$

that provides the upper bound to $P_{ic} \leq P_m + P_f$ according to the union bound [3].

Given an optimal design of ϕ matched with the observation model $p(\mathbf{y}|\mathbf{x}(m))$ and proper selection of N, one can demonstrate that P_f can be made negligibly small and does not depend on the number of items M in the database. Contrarily, P_{sa} and P_m depend on identification capacity $C_{id} = I(X;Y)$, where $I(.;.)$ denotes the mutual information [4] and can be made negligibly small, if the number of items in the database satisfies $\frac{1}{N} \log_2 M \leq C_{id}$. Moreover, it is important to point out that these probabilities of errors are equal, if all items in the database are generated to be independent and the attacker submits his probes independently to the data stored in the database [3]. Otherwise, these two terms should be investigated independently necessitating to analyze the behaviour of the informed attackers and the priors about the statistics of database [5]. Since this problem is out of scope of this paper, we will assume that the attacker is uninformed.

Therefore, in following, we will evaluate P_m and $P_f = P_{sa}$ for two different approaches in identification system design based on the synchronisation templates and on robust invariant features.

One of the main challenges in the optimal design of identification systems based on microstructure images consists in the lack of accurate statistical models $p(\mathbf{y}|\mathbf{x}(m))$ and $p(\mathbf{y}|\mathbf{x}')$ that describe the system performance under the hypothesis H_m and H_0, accordingly. The situation is complicated by the fact that the microstructure images are acquired at micrometer scale which requires a very accurate synchronisation mechanism. At this moment, up to our knowledge, there is no systematic comparison of different synchronisation mechanisms adapted to the micrometer images contrarily to computer vision applications where SIFT [6], SURF [7] and LBP-like features [8] gain more and more popularity. At the same time, it should be noticed that once the sets are perfectly synchronised one can approximate $p(\mathbf{y}|\mathbf{x}(m))$ and $p(\mathbf{y}|\mathbf{x}')$ very accurately by the Gaussian models with some correlated noise [9]. These results are possible due to the addition of special synchronisation templates or graphical symbologies used uniformly for all objects in the database.

In the case that the synchronisation templates or symbologies are absent or impossible to be added the only possibility is to use microstructures directly via an exhaustive search of all possibly geometrical transformations. Obviously, such an option is computationally prohibitively expensive and can only be used for small scale systems where M is small, for example, in the case of authentication, when the identity of the object is known, $M = 2$. Otherwise, invariant robust features might be deployed. However, early experiments have shown

that SIFT feature based registration on microstructures is by far not accurate enough for identification purposes. The fact however, that a geometric relation can be established between the features of a query microstructure and those of a database item, has proved to be a reliable identification heuristic, on which our second identification architecture is based.

Therefore, a fair comparison of template based and invariant features based methods represents an interesting research problem that will be considered in the following sections under the above introduced performance measures.

III. Template based identification

For the identification based on a pre-designed template we have used a set of points printed on the packages shown in Figure 3. The points are well detectable and are used to evaluate the parameters of an affine transform that models the mismatch between the images acquired at the enrolment and identification stages. Obviously, since the synchronisation template is the same for all objects under enrolment, the synchronisation can be performed without the explicit knowledge of data index m. The framework and procedure can be seen schematically in Figure 2a and 2b. Once the images are synchronised we use the additive model in the hypothesis testing [1]:

$$\begin{cases} H_0 & : \mathbf{y} = \mathbf{x}' + \mathbf{z}, \\ H_m & : \mathbf{y} = \mathbf{x}(m) + \mathbf{z}. \end{cases} \tag{5}$$

Following the assumption of the Gaussianity of the noise $\boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2 \mathbf{I}_N)$, one can show that the sufficient statistic for the identification problem formulated above, is the cross-correlation coefficient $\rho_{xy} = \mathbf{y}^T\mathbf{x} \geq \alpha N$ between vectors $\mathbf{x}$ and $\mathbf{y}$ with α to be a threshold [10]. It should be pointed out that the cross-correlation coefficient based decision rule $\rho_{x(m)y} \geq \alpha N$ is closely related to the Euclidian decision rule $\phi(\mathbf{y}, \mathbf{x}(m)) = \|\mathbf{y} - \mathbf{x}(m)\|^2 \leq \gamma N$ under the condition that $\|\mathbf{x}(m)\|^2$ are the same for all m.

The cross-correlation coefficient $\rho_{x(m)y}$ will be used in the following experimental tests on FAMOS dataset for 3 combinations of cameras to evaluate P_m^T and P_f^T:

$$P_m^T = \Pr[\rho_{x(m)y} \leq \alpha N | H_m], \tag{6}$$

$$P_f^T = \Pr[\cup_{m \neq m'} \rho_{x(m)y} \geq \alpha N | H_m]. \tag{7}$$

Figure 3: An example of the printed mark used for synchronisation in the template based identification framework.

IV. Invariant features based identification

In the case of robust invariant features, we will proceed with a set of SIFT descriptors [11] $\mathbf{x}^k(m)$, $1 \leq k \leq K$ extracted for each image $1 \leq m \leq M$, where K denotes the number of SIFT descriptors per image stored in the database. We also store the spacial location of all SIFT points. The SIFT descriptor is a 128-dimensional vector $\mathbf{x}^k(m) = (x_1^k(m), x_2^k(m), \cdots, x_{128}^k(m)) \in \mathbb{R}^{128+}$ representing histograms.

The probe data are represented by a set of descriptors $\mathbf{y}^j$, $1 \leq j \leq J$ extracted from the image $\mathbf{y}$ with $J \leq K$.

At this stage of the investigation, we ignore any complexity issues to evaluate the performance limits of the system. Obviously, carefully designed pruning and heuristics can help significantly to optimize the search or actual matching complexity. Therefore, in this paper we exhaustively perform a binary test for all images $\mathbf{x}(m)$, $1 \leq m \leq M$ to match the probe data with the data stored in the database. The test based on SIFT feature matching is formulated as $\phi^S(\mathbf{x}^k(m), \mathbf{y}^j) \leq \beta J$ with β stands for the threshold.

Since the complexity of the above rule might be prohibitively expensive even for the modern computers, we have used some obvious heuristics to constrain the search space. It should be pointed out that one is not interested to accurately estimate the parameters of desynchronization transform and then to compare the aligned images as in the template based identification. In principle features can be deployed to this end, but early experiments show that they perform significantly worse than the template based method. Contrarily, we will evaluate the co-occurrence statistics of SIFT descriptors in the probe image and images stored in the database. Therefore, given a probe image $\mathbf{y}$ and its descriptors $\mathbf{y}^j$, $1 \leq j \leq J$, it is reasonably to assume that (a) only non-ambiguously SIFT descriptors matched with the descriptors extracted from each image $\mathbf{x}(m)$ will be used for the co-occurrence estimation and (b) under the condition of correct matching the non-ambiguous matched descriptors should undergone the same geometrical transformation $\mathbf{A}$ that is assumed to be affine under the local approximation.

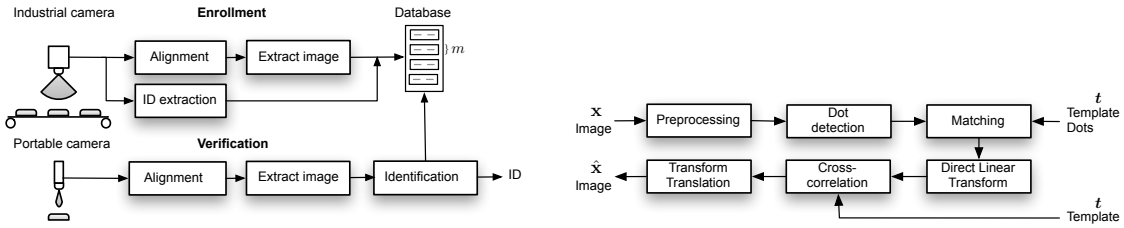
The SIFT based identification procedure, as shown in Figure 4a and 4b starts with the computation of Euclidian distance between the probe and database descriptors:

$$d^{jk}(m) = \|\mathbf{y}^j - \mathbf{x}^k(m)\|^2, \tag{8}$$

for all $1 \leq j \leq J$, $1 \leq k \leq K$ and $1 \leq m \leq M$. The SIFT features are assumed to be invariant to the geometrical transformations which only manifest itself in the addition of the Gaussian noise that can be taken into account by the Euclidian matching metric.

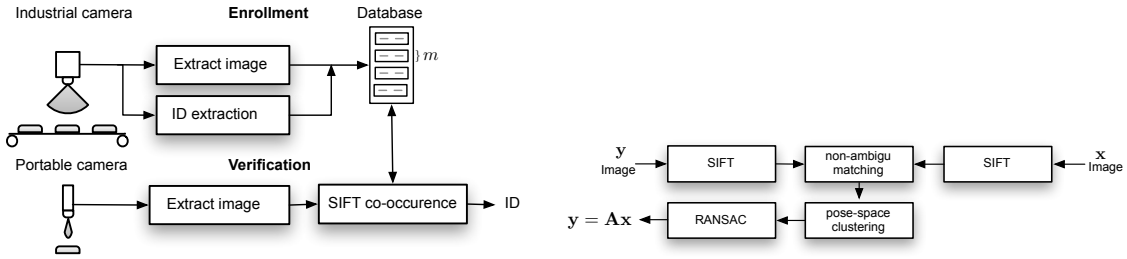
To prune the co-occurrence estimation, we use the observation (a) to select the unique matching pairs between the probe and image with the index m. Practically, we assume that if the match is unique, any remaining pair should produce a significantly lower matching score. For this purpose, the distances are ordered in the ascending order $d^{j(k)}(m)$ such that:

$$d^{j(1)}(m) = \min_{1 \leq k \leq K} d^{jk}(m), \tag{9}$$



(a) (b)

Figure 2: The acquisition and identification architecture for template based matching: 2a shows the enrollment and identification framework, 2b shows the synchronisation framework.



(a) (b)

Figure 4: The acquisition and identification architecture for SIFT based matching:4a shows the enrolment and SIFT co-occurrence identification framework, 4b shows the synchronisation framework.

and all pairs for which $d^{j(1)}(m) \geq 1.65 d^{j(2)}(m)$ are kept. Otherwise, the pair is rejected [6]. This results in $J'(m) \leq J$ matches per each index m.

Coarse hough pose-space clustering [12] is then deployed to further rid the set of outliers. At the next stage the RANSAC matching procedure [13] is applied to all non-ambiguous matches to estimate the parameters of local affine transform $\hat{\mathbf{A}}(m)$ for each index m under the condition (b). If such a unique transform is found the probe is declared as the object with the corresponding index $\hat{m}$ or otherwise it is rejected. An example of the end result of this procedure can be seen in Figure 5.

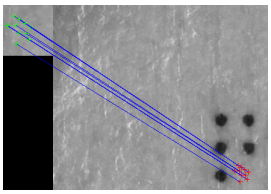


Figure 5: An example of Feature based matching between an enrolled sample and a different acquisition of an identical label. Note that here the dots serve no purpose in the matching procedure.

The identification performance based on the SIFT descriptor co-occurrence estimation is evaluated according to:

$$P_m^S = \Pr[\mathbf{y} \text{ is rejected } |H_m], \tag{10}$$

$$P_f^S = \Pr[\cup_{m \neq m'} \mathbf{y} \text{ is accepted as } m' | H_m], \tag{11}$$

V. Comparison on the FAMOS dataset

The template and SIFT based identification methods were tested on the FAMOS dataset [1]. Two colour cameras, designated Cam1 and Cam2 respectively, are deployed. Lighting in both cases consists of a white led ring light together with an angled one approximately 90mm above the surface. Cam1 has a resolution of 2592×1944 (5Mp) with a sensor size of 5.7×4.4mm and a pixel size of $2.2\mu m$. It has an optical magnification of $1:0.9$. Cam2 has a resolution of 1601×1201 (2Mp), a sensor size of 7×5.2mm, a pixel size of $4.4\mu m$ and no optical magnification. We will test three combinations of cameras used for the enrolment and identification, i.e., Cam1-Cam1, Cam2-Cam2 and Cam1-Cam2. The last setup corresponds to a mismatch in the resolution between cameras which is a realistic scenario in practice.

Figure 6 shows the probabilities P_f^T and P_m^T for the template based system for three sets of cameras. Notable, the lower resolution Cam2 outperforms the high resolution Cam1 and gives the most accurate identification performance. As expected, the most worse performance is observed when the enrolment and verification cameras are non-identical. This can be partly explained by the fact that sampling is needed to bridge the gap in resolution.

Table I shows P_f^S and P_m^S for the SIFT based system under the above combinations of cameras. Contrarily to the template based architecture, Cam1 achieves the best performance, not only in terms of the P_m^S, but images from Cam1 exhibit more numerous stable features, as seen in Table II. The setup where Cam2 is tested against Cam2 produces a P_m^S value that is twice as high as for Cam1 versus Cam1. Interesting, the mixed setup,

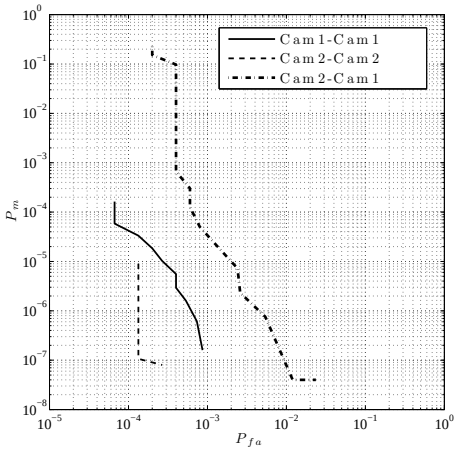


Figure 6: ROC curves for all cameras and template based matching.

does slightly better, most probably due to the fact the features from Cam1 images are of good quality.

	Cam1 vs Cam1	Cam2 vs Cam2	Cam2 vs Cam1
P_f^F	0	0	0
P_m^F	**0.040**	0.0773	0.0426

Table I: The probability of false alarm, P_f^S, and probability of miss P_m^S for feature based identification.

	Cam1 vs Cam1	Cam2 vs Cam2	Cam2 vs Cam1
Average number of inliers	**28.54**	5.26	5.86
Average percentage of inliers	**90.11**	59.82	58.74

Table II: The average number and percentage of inliers, or selected features, for the SIFT based identification system.

VI. Conclusions

In this paper, we considered physical object verification based on two techniques that extract features from synchronised images using inherently present graphical design marks and invariant features based on SIFT descriptors. Both techniques were compared on the FAMOS database set. Synchronisation based on a template mark gives stable performance and runs with relatively little parameter tuning. Because samples are rid of all geometric disturbances, this architecture has great potential for the application of further hashing and binarization schemes as is also common for human biometrics. The only drawback is the fact that a suitable mark must be present, or needs to be added, a measure that is unrealistic and costly for a lot of physical goods. SIFT based matching based on features in the microstructures themselves perform reasonably well on the FAMOS dataset and doesn't require any modifications to the objects. They are however, much more volatile in their performance and critically depend on excellent lighting and acquisition conditions. Further more, the identification procedure requires an exhaustive search over all candidates and is therefore slow.

Further work will concentrate in two primary directions. First, further investigation in to the statical properties of the FAMOS microstructures. Secondly, we will continue working on designing features specifically tailored towards the little patterns the microstructures exhibit.

All code, figures and the FAMOS dataset can be accessed directly from `http://sip.unige.ch/famos`.

VII. Acknowledgments

This work is supported by SNF-grant 200020-146379.

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] F. Farhadzadeh, S. Voloshynovskiy, O. Koval, F. Beekhof, and T. Holotyak, "Information-theoretical analysis of content based identification for correlated data," in *IEEE Information Theory Workshop, ITW2011*, Dublin, Ireland, Aug.30-Sep.3 2011.

[3] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *IEEE Information Theory Workshop, ITW2010*, Dublin, Ireland, Aug.30-Sep.3 2010.

[4] F. Willems, "J.p.: On the capacity of a biometrical identification system," in *In: Proc. of the 2003 IEEE Int. Symp. on Inf. Theory*, 2003, pp. 8–2.

[5] F. Beekhof, S. Voloshynovskiy, and F. Farhadzadeh, "Content authentication and identification under informed attacks," in *Proceedings of IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5 2012.

[6] M. Brown and D. Lowe, "Invariant features from interest point groups," 2002. [Online]. Available: citeseer.ist.psu.edu/brown02invariant.html

[7] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *In ECCV*, 2006, pp. 404–417.

[8] P. M., "Image analysis with local binary patterns." In: Image Analysis, SCIA 2005 Proceedings, Lecture Notes in Computer Science 3540, Springer, 115-118, plenary presentation, 2005.

[9] F. Beekhof, S. Voloshynovskiy, M. Diephuis, and F. Farhadzadeh, "Physical object authentication with correlated camera noise." in *15th GI-Symposium Database Systems for Business, Technology and Web*, march 2013.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory (v. 1)*, 1st ed. Prentice Hall, Apr. 1993.

[11] A. Vedaldi and B. Fulkerson, "Vlfeat: An open and portable library of computer vision algorithms," 2008. [Online]. Available: http://www.vlfeat.org

[12] C. F. Olson, "Efficient pose clustering using a randomized algorithm," *International Journal of Computer Vision*, vol. 23, no. 2, pp. 131–147, 1997.

[13] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.