 DOI: 10.3217/978-3-85125-603-1-21

Semi-Automatic Retrieval of Toolmark Images

Manuel Keglevic[1] and Robert Sablatnig[1]

***Abstract*— In order to identify and solve connected cases forensic experts are currently comparing toolmarks from crime scenes manually. However, especially for frequently occurring crimes like burglaries this task is cumbersome. In order to support the work of the forensic experts, we propose a semi-automated system for finding similarities between toolmark images in large databases. Our methodology uses convolutional neural networks to compute local image similarities in these toolmark images. This work presents the proposed approach and the evaluation conducted on a dataset of more than 3,000 toolmark images collected from real criminal cases.**

I. Introduction

A technique commonly used for break-ins by criminals all over Europe is the so-called *lock snapping*. Using a tool, like for instance a locking plier, door locks can be quietly broken (*snapped*) in a short amount of time using force and leverage. This, however, leaves unique imprints, i.e. toolmarks, of the pliers used on the cylinder locks. As an example, Figure 1 shows multiple toolmarks of the same tool on a broken lock cylinder. By comparing two different toolmarks using a comparison microscope forensic experts can assess if the marks were made by the same tool. This can either be used to confirm that a seized tool was used to commit a crime, or to link multiple cases together and thereby significantly support the investigation of such offenses. Furthermore, the toolmarks found on these locks are crucial as evidence in the following court cases.

Yet, the manual examination and comparison of the toolmarks found is a cumbersome task due to the number of burglaries occurring every year. Therefore, within the project FORMS we developed a semi-automatic system in order to assist the forensic experts. The proposed system consists of an application, which enables the forensic experts to catalog and search toolmark images in a central database, and a methodology based on machine learning. This methodology computes similarities between toolmark images automatically in regions manually annotated by the forensic experts. This way, the forensic experts are presented with a list of toolmarks sorted by similarity in order to reduce the amount of images requiring manual examination. The project FORMS was conducted in cooperation with the Bundeskriminalamt (Criminal Intelligence Service Austria), the CogVis GmbH and VICESSE. It started in Fall 2015 and concluded in February 2018 and was funded by the Austrian Security Research Programme KIRAS.

This paper is structured as follows: firstly, the state of the art in automatic toolmark comparison is presented. Secondly, the dataset, which was created in cooperation with the austrian police, is describe. Thirdly, the methodology based on learning local images similarities using Convolutional Neueral Networks (CNNs) and its evaluation is presented. Finally, we conclude with the advantages and disadvantages of our proposed system and an outlook for future work beyond the FORMS project.

[1]Computer Vision Lab, TU Wien, 1040 Vienna, Austria keglevic@cvl.tuwien.ac.at

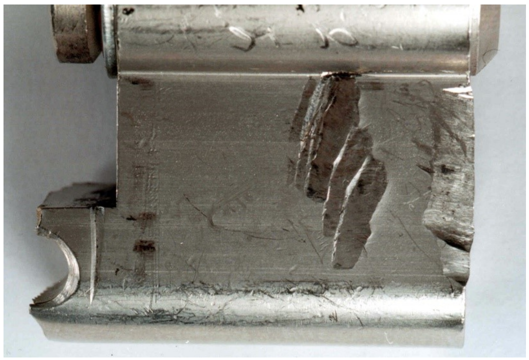

Fig. 1: Snapped lock with multiple toolmarks.

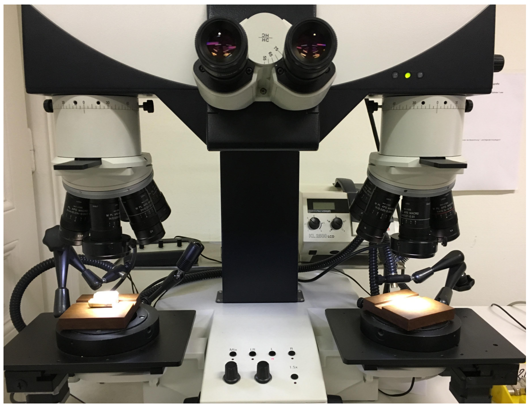

Fig. 2: Leica comparision microscope used by the forensic experts in Austria.

II. State of The Art

Since the validity of comparative forensic examination of toolmarks has been challenged in court, the development of automatic tools for the comparative examination of toolmarks has been in focus of the forensic community to obtain statistical support for the notion of the *uniqueness* of toolmark patterns [14], i.e. the existence of ‘measurable feature with high degree of individuality” [1]. For the comparison of striated toolmarks this led to a variety of methodologies [1], [2], [4], [3], [7], [8], [12] which operate on 1D profiles extracted from either 2D images or 3D surface

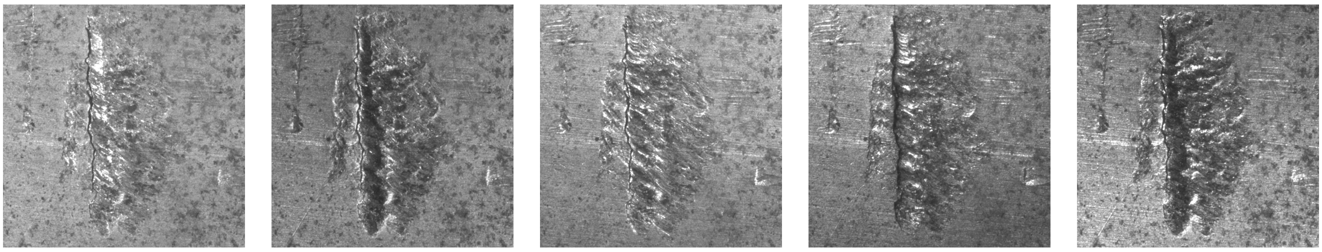

Fig. 3: Toolmark crop under different lighting conditions.

	Tools	Locks	Sides	Images
2015	25	115	230	1,782
2016	23	82	164	1,263
total	48	197	394	3,046

TABLE I: Statistics of the captured toolmark images divided by year.

scans. Similarity scores are commonly computed using either global [2], [4], [3] or local [8] Cross-Correlation (CC). Bachrach et al. [1] propose the use of locally normalized squared distances as similarity measure. In contrast to these approaches, Petraco et al. [12] propose an approach based on machine learning by using dimensionality reduction and Support Vector Machines (SVM) for the classification of the tool. More recently it was shown that CNNs outperform other methods by learning a similarity measure for striated toolmarks [11], [13].

However, all those experiments were carried out on striated toolmarks created under laboratory conditions. This includes for instance fixed angles of attack, constrained lighting conditions, high resolution 3D surface scans and hand-selected tools and surface materials as shown for instance by the only publicly available dataset; the NFI Toolmark dataset created by Baiker et al. [2].

III. Dataset

In order to allow an evaluation of the performance under real-world conditions, we created a dataset using lock cylinders. These lock cylinders were seized by the austrian Police in the course of break-in investigations in Vienna during the years 2015 and 2016. All images were captured using the Leica comparison microscope used by the forensic experts, which is shown in Figure 2. In order to allow an evaluation of the influence of different lighting conditions, a variable light ring was utilized to capture the toolmarks under 11 different lighting conditions. In Figure 3 the effect of different lighting conditions is shown on an example. In total, 197 lock cylinders from 48 linked cases were photographed on both sides. The resulting 3,046 images were divided into training set and test set by year, i.e. 2015 for training and 2016 for testing. In Table I the number of images and locks for each set is listed.

To provide matching local image similarities, matching patches in the images were annotated using a plugin developed for the image viewer nomacs. In this tool polylines are used to describe the toolmark edges and matching toolmarks are fitted using transformation matrices. By utilizing the fact that the cylinder locks are an approximately flat surface, the capturing angle is orthogonal to this surface and the distance of the camera is always the same, the possible transformations can be restricted to translations and rotations. In this way, matching patches can simply be extracted by moving a window along the polylines and their matching (transformed) counterparts. Matching toolmarks can either be found on the same lock cylinder, as shown in Figure 4, or on different lock cylinders from the same linked case. Since all cylinder locks photographed originate from linked cases, it is guaranteed that for each tool multiple toolmarks exist.

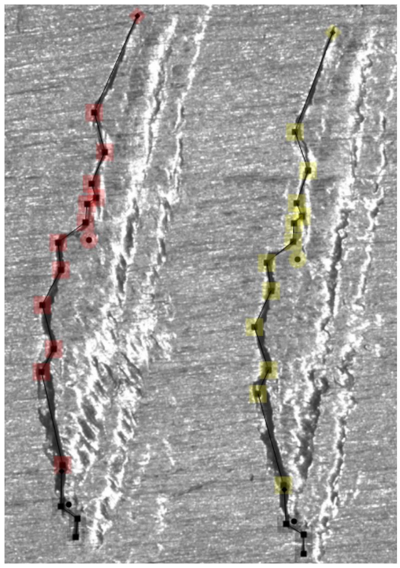

Fig. 4: Matching toolmarks annotated using a polyline and fitting with a transformation matrix.

To allow for training and evaluation of different local image similarity approaches, 41,030 and 25,014 images patches were extracted from the training set and the test set, respectively. Additionally, 50,000 matching and 50,000 non-matching image pairs were created to enable comparable evaluations.

IV. Methodology

In this section, both parts of the proposed methodology are decribed. Firstly, the neural network used to compute local images similarities is shown. Secondly, our two approaches to combine the local image similarities for the retrieval of similar toolmark images are presented.

A. Local Image Similarities

Our proposed neural network with triplet loss is based on the work of Balntas et al. [5]. Similar to siamese networks [6] the network architecture consists of multiple branches with shared weights as shown in Figure 5.

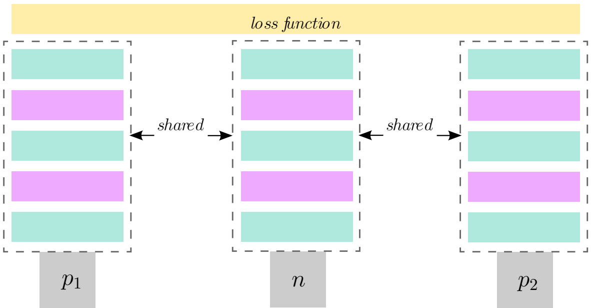

Fig. 5: Triplet architecture

The training is performed by forwarding three input samples, i.e. a triplet, through these equal CNN branches. Each triplet consists of an anchor x_{p_1}, a positive (matching) sample x_{p_2} and a negative (non-matching) sample x_n. The loss function then combines the three outputs $f(x_i)$ and the error is back-propagated. In contrast to other triplet loss functions, like the SoftMax Ratio proposed by Hoffer et al. [9] which only takes one negative distance into account, all three distances between the samples are used:

$$\begin{aligned} \Delta^+ &= \left\|f(x_{p_1}) - f(x_{p_2})\right\|_2 \\ \Delta_1^- &= \left\|f(x_{p_1}) - f(x_n)\right\|_2 \\ \Delta_2^- &= \left\|f(x_{p_2}) - f(x_n)\right\|_2 \end{aligned} \tag{1}$$

Instead of forcing the distance Δ^+ to be just smaller than Δ_1^-, it is forced to be smaller than $\Delta^* = min(\Delta_1^-, \Delta_2^-)$. In this way negative mining is performed implicitly as illustrated in Figure 6.

The loss function is then defined as:

$$\ell(T) = \left(\frac{e^{\Delta^+}}{e^{\Delta^+} + e^{\Delta^*}}\right)^2 + \left(1 - \frac{e^{\Delta^*}}{e^{\Delta^+} + e^{\Delta^*}}\right)^2 \tag{2}$$

In contrast to the network proposed by Balntas et al., we employ a DenseNet CNN architecture for each of the branches [10].

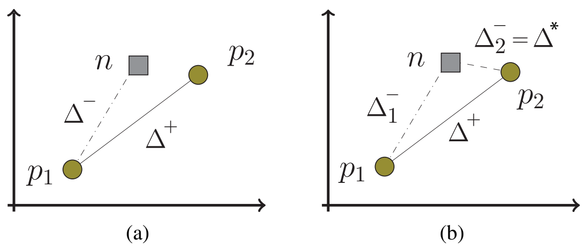

Fig. 6: SoftMax Ratio (a) compared to SoftPN (b) [5].

B. Toolmark Retrieval

In order to retrieve toolmark images, the local image similarities have to be combined to form a similarity measure. For this we use two different approaches.

Firstly, local patches are extracted along the annotated toolmark edges in fixed steps and their features are computed using the neural network described in the previous section. The features are then pairwisely compared using the euclidean distance for each step. The resulting distance between two toolmarks is then computed by summing up these distances and normalizing by length. For toolmarks with different length, the alignment is shifted until a minimal distance is found. The advantage of this approach is, that it is simple and computationally inexpensive. However, it requires an exact annotations since otherwise patches on different parts of the toolmark may be compared against each other. Small variations can lead to accumulated length differences which cannot be compensated by this approach as shown in Figure 7.

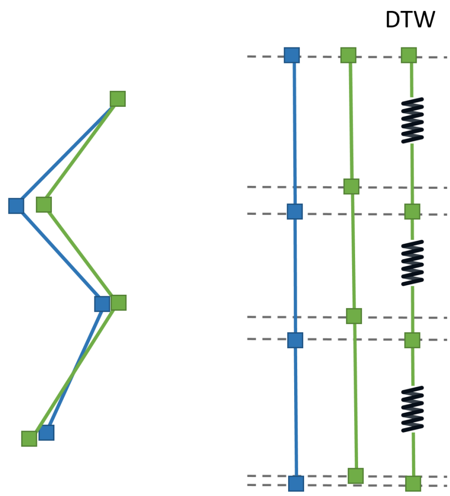

Fig. 7: Matching in fixed step sizes compared to a distance computation using dynamic time warping (DTW).

Secondly, to relax the requirement of a fixed step size, dynamic time warping (DTW) is proposed to allow for a more flexible matching of the local images patches. This way small inaccuracies in the annoation process and the resulting changes in the length of the toolmark segments can be compensated. In Figures 7 on the right the advantage of the DTW approach is visualized.

V. Evaluation

For computing the local image similarities the neural network is trained using the extracted image patches of the training set and evaluated on the test set. Two different approaches were evaluated for the selection of the positive samples.

The first strategy was to define positive samples as matching patches from different locks and different lighting conditions in order to train the network to the high variability of materials and lighting conditions. However, using this strategy a false positive rate at 95% recall (FPR95) of only about 80% percent is achieved on 100,000 evenly distributed matching and non-matching pairs of patches. This can be explained by the high variability of the presented image patches and human errors in the annoation process.

In order remove the influence of human errors in the annoation process and restrict the variability of the matching patches, as a second strategy, positive samples were defined as patches from just different lighting directions on exactly the same position. This way, an FPR of under 30% is achieved on 100,000 matching and non-matching images pairs which were selected using the same strategy. One interpretation for the remaining false positives is that many patches, mainly from locks made out of shiny materials, are indistinguishable due to the limited dynamic range of the images.

Using this trained network to compute the local image similarities, a cumulative match score of about 70% at a retrieval rate of 20% can be achieved for the toolmark images. In Figure 8 the cumulative match characteristic is depicted for both the approach using a fixed step size and the DTW method. It can be seen, that the fixed approach performs slightly better since in this case the annotations were done precisely. Yet, the DTW approach provides comparable results with the advantage of an added flexibility in the annotation process.

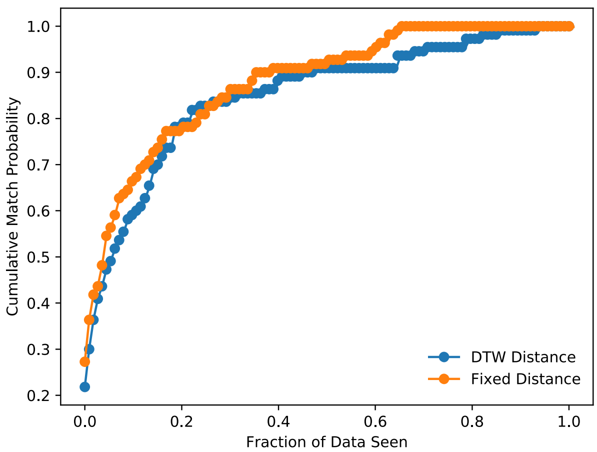

Fig. 8: Cumulative match characteristic on the test set using either a fixed step size or dynamic time warping. The data is retrieved ranked by similarity.

VI. Conclusion

In this paper a two step approach for computing toolmark similarities was presented. Firstly, a neural network using a triplet architecture was proposed to compute local image similarities. Secondly, two approaches for combining local image similarities to form distance scores for toolmark images were shown. The proposed system was evaluated on a toolmark dataset created by photographing cylinder locks from real criminal cases. It was shown that with a probability of more than 70% a matching toolmark is found in case 20% of the images in a database are retrieved. Even though this leaves room for improvement, these results are promising and show that an automated retrieval systems can valuably support the work of forensic experts. For future work, the proposed approaches will be extended to other areas of forensic images; as for instance footwear impressions.

Acknowledgment

This work has been funded by the Austrian security research programme KIRAS of the Federal Ministry for Transport, Innovation and Technology (bmvit) under Grant 850193. We would like to thank the forensic experts of the Criminal Intelligence Service Austria and the LKA Wien (AB08 KPU) for their help. The Titan X used for this research was donated by the NVIDIA Corporation.

References

[1] B. Bachrach, A. Jain, S. Jung, and R. D. Koons, "A Statistical Validation of the Individuality and Repeatability of Striated Tool Marks: Screwdrivers and Tongue and Groove Pliers," *Journal of Forensic Sciences*, vol. 55, no. 2, pp. 348–357, 2010.

[2] M. Baiker, I. Keereweer, R. Pieterman, E. Vermeij, J. van der Weerd, and P. Zoon, "Quantitative comparison of striated toolmarks," *Forensic Science International*, vol. 242, pp. 186–199, 2014.

[3] M. Baiker, N. D. Petraco, C. Gambino, R. Pieterman, P. Shenkin, and P. Zoon, "Virtual and simulated striated toolmarks for forensic applications," *Forensic Science International*, vol. 261, pp. 43–52, 2016.

[4] M. Baiker, R. Pieterman, and P. Zoon, "Toolmark variability and quality depending on the fundamental parameters: Angle of attack, toolmark depth and substrate material," *Forensic Science International*, vol. 251, pp. 40–49, 2015.

[5] V. Balntas, E. Johns, L. Tang, and K. Mikolajczyk, "PN-Net: Conjoined Triple Deep Network for Learning Local Image Descriptors," *ArXiv*, 2016.

[6] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a Similarity Metric Discriminatively, with Application to Face Verification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 539–546.

[7] W. Chu, R. M. Thompson, J. Song, and T. V. Vorburger, "Automatic identification of bullet signatures based on consecutive matching striae (CMS) criteria," *Forensic Science International*, vol. 231, no. 1, pp. 137–141, 2013.

[8] L. S. Chumbley, M. D. Morris, M. J. Kreiser, C. Fisher, J. Craft, L. J. Genalo, S. Davis, D. Faden, and J. Kidd, "Validation of tool mark comparisons obtained using a quantitative, comparative, statistical algorithm." *Journal of Forensic Sciences*, vol. 55, no. 4, pp. 953–61, 2010.

[9] E. Hoffer and N. Ailon, "Deep metric learning using Triplet network," *ArXiv*, 2014.

[10] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *CoRR*, vol. abs/1608.06993, 2016. [Online]. Available: http://arxiv.org/abs/1608.06993

[11] M. Keglevic and R. Sablatnig, "Learning a Similarity Measure for Striated Toolmarks using Convolutional Neural Networks," in *Proceedings of the 7th IET International Conference on Imaging for Crime Detection and Prevention (ICDP)*, 2016.

[12] N. D. K. Petraco, H. Chan, P. R. D. Forest, P. Diaczuk, C. Gambino, J. Hamby, F. L. Kammerman, W. Brooke, T. A. Kubic, L. Kuo, G. Petillo, E. W. Phelps, A. Pizzola, and D. K. Purcell, "Application of Machine Learning to Toolmarks - Statistically Based Methods for Impression Pattern Comparisons," NCJRS (239048), Tech. Rep., 2012.

[13] R. Sablatnig, "Retrieval of striated toolmarks using convolutional neural networks," *IET Computer Vision*, 2017. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/iet-cvi.2017.0161

[14] R. Spotts, L. S. Chumbley, L. Ekstrand, S. Zhang, and J. Kreiser, "Optimization of a Statistical Algorithm for Objective Comparison of Toolmarks," *Journal of Forensic Sciences*, vol. 60, no. 2, pp. 303–314, 2015.