

Airport Cable-cutting Toolmark Rapid Tracing Based on Single-point Laser Sensing

Nan Pan,^{1*} Gang Li,² Dilin Pan,³ Xin Shen,⁴ and Xin Wang⁴

¹Faculty of Civil Aviation and Aeronautical, Kunming University of Science & Technology,
No. 727 Jingming South Road, Chenggong University Town, Kunming 650500, P.R. China

²Institute of Forensic Science, Shijiazhuang Public Security Bureau,
No. 66 Yuannan Road, Qiaoxi District, Shijiazhuang, Hebei 050021, P.R. China

³Kunming SNLab Technology Co., Ltd.,

Architect Business Center 2202 Room Dong Feng Dong Road 47#, Kunming 650228, P.R. China

⁴Metrology Center of Yunnan Power Grid Co., Ltd., No. 73 Tuodong Road, Kunming 650017, P.R. China

(Received April 16, 2018; accepted April 15, 2019)

Keywords: airport cable, laser displacement sensor, toolmarks, threshold sequence, difference in variance

Because current image and 3D scanning file processing technologies are difficult to use in comparing cable-cutting toolmarks effectively, an airport cable-cutting toolmark rapid-tracing matching algorithm based on single-point laser sensing is presented. The proposed algorithm applies a boxplot to the linear toolmark signals picked by a laser displacement sensor to perform an abnormal data correction. Then, a rotation angle adaptive correction is performed to unify the matching data. Furthermore, in multimatching strategies based on a threshold sequence, the difference in variance is used to perform the similarity matching of toolmark features. Finally, a correct determination of the corresponding tools is rapidly achieved. The practicality and effectiveness of the proposed algorithm are verified by experiments using actual mark inference cable-cutting tools.

1. Introduction

In recent years, frequent theft of cables in airports has resulted in huge state property losses and caused interruption in communication signals and equipment power supply. This has caused failure in the respective systems, leading to several accidents, significant loss of life, and diminished safety of property.⁽¹⁾ Criminals often use wire cutters, cable cutters, pliers, and other large cutting tools to sever cables. Toolmarks, which are scratches on the surface of the body, could be caused by the pressure of the line-shaped deformation; these toolmarks, in the form of broken ends on the surface, are frequently found at the scene. They exhibit characteristics that are difficult to destroy or disguise, occur frequently, and have high identification values for investigators to determine the nature of the case and the tools used in the criminal act. These characteristics are crucial for identifying suspects.⁽²⁾

Traditionally, two toolmarks are compared by a toolmark examiner using a comparison microscope. Two toolmarks are put side by side, and striations are accentuated using an oblique

*Corresponding author: e-mail: 15808867407@163.com

<https://doi.org/10.18494/SAM.2019.2269>

light, thus illuminating ridges and shading furrows.⁽³⁾ The result is a light-shadow pattern, representing the actual toolmark topography. The expert then performs a visual comparison of the two illumination patterns and attempts to identify matching striations. The goal is to determine whether striation patterns are in sufficient agreement or disagreement to conclude whether they are of common or different origins, respectively.

2. Signal Preprocessing

2.1 Abnormal data correction

Most of the abnormal data in this analysis are caused by excessive reflection, in which there is a significant difference between the abnormal data point and the data nearby.⁽⁴⁻⁶⁾ Data that can be identified as abnormal are data that do not conform to the established trend, that is, they are very low or very high. Overall, the abnormality is determined by the following two rules:

- 1) For continuous sampling data, the magnitude of the change must be limited to a certain range. Assuming that the range is C , C is the maximum above all the sampling data with a certain probability of occurrence, and there are many data near its value.
- 2) The rate of change in the data does not appear to be different from the usual growth rate. If the slope of the waveform varies with the surroundings, abnormal data are obtained.

Thus, if any of the above occurs, data are considered abnormal, and when abnormal data are present, a normal data correction should be performed depending on the nearby normal data. The boxplot is a type of pattern containing straight lines and boxes, which can directly reflect the sample distribution trend data. The boxplot mainly consists of five parts, namely the minimum Min, first quartile Q1, median M, third quartile Q3, and maximum Max.

If the probability $0 < p < 1$, the quantile Z_a of the random variable X or its probability distribution is defined, and the real number satisfying the condition $P(X > Z_a) = a$ is the quantile.

Mapping based on the above five statistics:

- 1) Draw an axis with a consistent unit of measure for batch size and data units; the starting point is slightly smaller than the minimum, and the length of the batch is smaller than that of full-pitch data.
- 2) Draw a rectangular box; where the locations of the two ends of the data correspond to the upper and lower quartiles (Q1 and Q3). Draw a line segment as the median line in the median (X_m) position inside the rectangular box.

The above is the boxplot. When the observed data in the dataset is unusually larger or smaller than the other data in the dataset, the observation value will cause the maximum or minimum end of the box graph to be anomalously away from the middle box, for which a set of rules is established to amend the problem. The amendment process is as follows:

- 1) The first quartile is known as Q1, the third quartile is Q3, and the distance between them is $IQR = Q3 - Q1$, which becomes the quartile pitch.
- 2) All data less than $Q1 - 1.5IQR$ and greater than $Q3 + 1.5IQR$ are marked as abnormal data and are individually marked as abnormal points when drawing, which is no longer in the body of the boxplot.

With this correction step, most of the abnormal values are filtered from the statistical point of view.

2.2 Data rotation correction

In the engineering field, the inclination refers to the ratio of the settlement difference between the two ends of the base and the distance. According to the actual situation, assuming that the toolmark is given a signal of length n , where mid is the midpoint, the tilt can be defined as

$$RotateRange = \frac{\sum_{mid+1}^n S_i - \sum_1^{mid-1} S_i}{mid} \begin{cases} < 0, \text{ Tilt to the right} \\ = 0, \text{ Balanced state} \\ > 0, \text{ Tilt to the left} \end{cases} \quad (1)$$

Each toolmark detection signal is generally required to perform the appropriate rotation correction, performing different amplitude corrections according to the different positions, on the basis of the known *RotateRange*, wherein, for each point

$$Rnews_i \leftarrow s_i + RotateRange \times \frac{mid - i}{mid - 1} \quad (2)$$

3. Similarity Matching

3.1 Variable length and partial overlap problem

The most common way to similarity-match two signals is to calculate the difference between these two signals and accumulate all the differences.^(7,8) The larger the final result is, the greater the degree of deviation is and the lower the similarity is. Considering the existence of errors in the actual detection, small differences can generally be ignored; a threshold could be added on this basis, with two sections of the signal in a certain position within a certain range considered as equal.

Prior to the similarity matching of toolmark signals after noise reduction, the following two issues must be addressed:

- 1) Variable toolmark length. Toolmark detection signal lengths are not the same. Most of the lengths of the matching signal data and the signal to be matched are different. In this case, the similarity of two discrete sequences could not be directly measured using the Euclidean distance and correlation coefficients. Thus, point-to-point operations become meaningless.
- 2) Parts overlap. This means that two detection signal marks may only overlap at a certain part by coincidence. This condition can cause significant interference with the calculation of the final coincidence.

Therefore, the problem of variable length and overlap can be optimized through a computing algorithm capable of matching. The basic steps are as follows:

- 1) Set the input data of A and B, which are data that satisfy the above requirements.
- 2) Set a match to the minimum longest L. The two coincidences must meet the minimum overlap length by selecting the largest length to the shortest part from A and to compare it with that from B, that is, equivalent to choosing a different location for a multitude of matches.
- 3) Iteratively execute the contrast for each position. Each comparison should be compared with the variance of the degrees of differences of the two corresponding positions. The current state is recorded if the variance is minimum.
- 4) If the function of 3) is completed, the roles of A and B are exchanged, followed by the completion of steps 2) and 3).
- 5) Calculate the variance of the minimum difference degree and output the matching result.

Multithreaded programs can be assigned to different CPU cores; this is a simple and practical way of parallel design. However, they also face a series of problems such as multithread scheduling, resource sharing, and shared lock. Combined with actual usage scenarios, using a more coarse-grained, multithreaded mode of operation could prevent resource sharing and shared lock, avoiding the necessity of much multithreaded code design. In this manner, efficient algorithm libraries could be loaded directly.

By combining a test toolmark with a sample toolmark in the sample pool as a task, each test sample was randomly placed in a thread pool. The number of thread pools and the number of concurrent computations take full advantage of the current CPU core number. Each task performs a different matching strategy; the calculation of each task will be merged in subsequent steps.

3.2 Matching strategy based on threshold sequence

The most common way to similarity-match two signals is to calculate their difference and accumulate all the differences. The larger the final result is, the greater the degree of deviation is and the lower the similarity is. Considering the existence of errors in the actual detection, small differences can generally be ignored; a threshold could be added on this basis, with two sections of the signal in a certain position within a certain range considered as equal.

The similarity based on the threshold difference can be calculated in this manner under different variations. It is possible to find a transformation of the least degree of difference within a given transformation range.

At this time, it is assumed that the mark detection signals A and B are intercepted, and the converted inputs are $I_1 = \{i_{11}, i_{12}, \dots, i_{1m}\}$ and $I_2 = \{i_{21}, i_{22}, \dots, i_{2m}\}$.

The degree of difference is calculated as

$$Difference = \frac{\sum_{k=1}^m g(i_{1k} - i_{2k}) \times context(k)}{m}, \quad (3)$$

$$g(x) = \begin{cases} 0, & |x| \leq c \\ cost(x), & |x| > c \end{cases}, \quad (4)$$

where c is the given threshold, $cost(x)$ is a cost function, and $cost(x) > 0$. The parameter $context(k)$ is a combination of the previous match and the weight of the situation, which is mainly to consider the position k before the match.

Considering the difference in the value of each of the two inputs, if the difference between the two values is within a certain range, then it is within the threshold, regardless of the difference. If the difference is greater than the threshold, then calculate the difference based on the cost function $cost(x)$ and add it to the final result. If the result is 0, then it is exactly the same. The larger the value is, the greater the difference is.

3.3 Matching strategy based on difference in variance

Variance is a measure of the degree of dispersion when measuring a random variable or a set of data. There is some work that needs to be carried out first before applying it to signals:

Assuming two input signals, $I_1 = \{i_{11}, i_{12}, \dots, i_{1m}\}$ and $I_2 = \{i_{21}, i_{22}, \dots, i_{2m}\}$, calculate the absolute value of the difference between the two signals and then calculate the variance

$$var(g(I_1 - I_2)^v), \quad (5)$$

where v is the difference in power, the value of which is generally 2, and $g(x)$ is the mapping function. To prevent the outliers or individual data points from interfering with the final result, the function $g(x)$ of the logarithmic function or other nonlinear forms can be chosen for the mapping of the data. Generally, when the difference is greater than a certain value, the overall result of the dot interference is reduced. The result of Eq. (3) is the degree of deviation between the two input signals. If the difference is constant, the two signals are proven to be exactly the same in shape, and if the difference is very different, the two signals are proven to be different.

4. Experimental Testing

The effectiveness of this algorithm is verified through the actual cutting tool source experiment. The experimental setup is as follows:

Three tools that are usually employed in airport cable theft cases were selected: wire cutters (A), pliers (B), and steel wire clamps (C). A copper bar of 1 cm diameter was cut by cutting breakage. All breakage surfaces were tested using the toolmark single-point laser detection equipment; its specific parameters are shown in Table 1. The sample parameter setting involved the following: a laser spot diameter of 1.25 μm , a subdivided figure for 3200 steps/s, a sample pulse frequency of 1000 Hz, a sampling interval of 50 ms, and a sampling frequency of 20 Hz, the sampling points being determined according to the cross-sectional area of the broken end. The related algorithms that match the program were coded in Python upon verification by Matlab 2018a. The program was run on a PC with an Intel Core i7 4.2GHz CPU with 16G DDR4 memory (Fig. 1).

The 30 sets of data labeled T1–T30 (T10 and T11 being substantially the same) were used as test data. The sample library contained 1000 datasets containing data on cutting by 10 different



Fig. 1. (Color online) Practical testing environment.

Table 1
Sample matching results obtained by the algorithm proposed in this study.

Data number	Highest matching data	Accurate traceability rate (%)	Data description
T1	T2	100	Group A benchmark data
T2	T1	100	Group A benchmark data duplicate detection
T3	T5	100	Group A benchmark data displacement 1/6
T4	T5	100	Group A benchmark data displacement 1/3
T5	T6	100	Group A benchmark data displacement 1/3
T6	T5	100	Group A benchmark data displacement 1/2
T7	T8	100	Another line with the some broken end A
T8	T3	100	T7 based on the direction D of translation of 1/6 of the other line
T9	T14	20	T7 based on the direction U of translation of 1/6 of the other line
T10	T26	20	Dislocation 1/3 and deletion of 2/3 based on T7
T11	T12	100	Group B benchmark data
T12	T11	100	Group B benchmark data duplicate detection
T13	T12	100	Group B benchmark data duplicate detection
T14	T15	100	Group B based on the direction D of translation of 1/10 of the other line
T15	T14	100	T14 duplicate detection
T16	T17	100	Group B based on the direction U of translation of 1/10 of the other line
T17	T16	100	T16 duplicate detection
T18	T19	100	Group B similar tools, different ends
T19	T18	100	T18 duplicate detection
T20	T21	100	T18 based on the direction D of translation of 1/10 of the other line
T21	T20	100	T20 duplicate detection
T22	T23	80	T20 based on the direction D of translation of 1/10 of the other line
T23	T22	80	T23 duplicate detection
T24	T25	80	Group C benchmark data
T25	T24	80	Group C benchmark data duplicate detection
T26	T10	60	Group C based on the direction D of translation of 1/10 of the other line
T27	T28	80	T26 duplicate detection
T28	T27	80	T26 duplicate detection
T29	T30	80	Group C based on the direction U of translation of 1/10 of the other line
T30	T17	100	T29 duplicate detection

tools, which often appear in airport cable theft cases. Excluding the data in the sample library, the results show the first five top ranking values.

There were three sets of data in T1–T30. T1–T10 were from the marks formed by tool A, T11–T23 were from the marks formed by tools B1 and B2, and T24–T30 were from the marks formed by tool C. B1 and B2 are two tools that belong to the same tool group B. To make the simulation of the data acquisition in the crime scene more realistic, each group of test data was required to be tested again after shifting the position based on the benchmark toolmark data and form new data. The data in A mainly contained lateral displacements, that is, data moving in a straight line from the original marks. Some data coincided with original data after the movement. At the same time, all the data in A, B, and C had U-direction (up) and D-direction (down) movements, and a certain degree of dislocation with the original benchmark marks.

As shown in Fig. 2, T16 (blue) and T18 (red) are repetitive detection signals of the same toolmark. The difference in the case of the smallest difference between the two datasets is shown in the third column. The overlay of the two overlapped signals is shown in the fourth column when the match is completed. Visual observation shows that the signals have a high degree of overlap. During detection, some errors would inevitably be produced as the two cannot exactly be the same. The test results show that the similarity is 90%, which means that the two overlapped signals have been effectively aligned during matching. The calculation takes 12 s, with a matching success rate of 90% and a failure rate of 10% .

For comparison with the technology proposed in this study, the algorithm from Ref. 8 is applied to the matching test using the same 30 sets of data labeled T1–T30. The results are

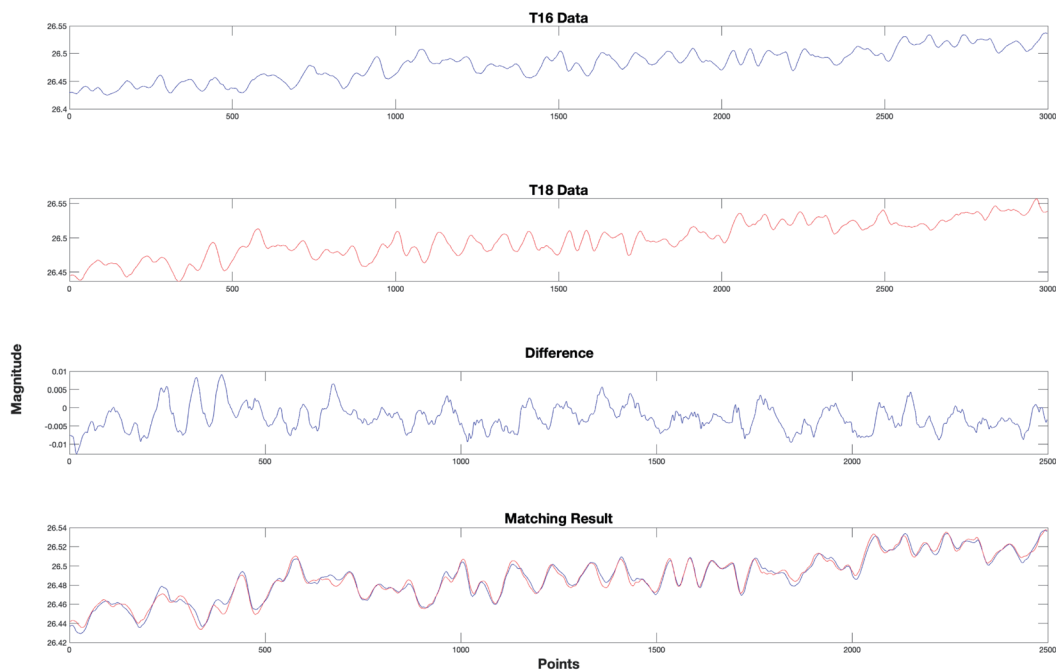


Fig. 2. (Color online) Signal difference matching.

Table 2
Sample matching results obtained by the algorithm published in Ref. 8.

Data number	Highest matching data	Accurate traceability rate (%)	Data description
T1	T2	80	Group A benchmark data
T2	T1	60	Group A benchmark data duplicate detection
T3	else	20	Group A benchmark data displacement 1/6
T4	else	20	Group A benchmark data displacement 1/3
T5	else	20	Group A benchmark data displacement 1/3
T6	T1	60	Group A benchmark data displacement 1/2
T7	T8	60	Another line with the broken end A
T8	T7	80	T7 based on the direction D of translation of 1/6 of the other line
T9	else	0	T7 based on the direction U of translation of 1/6 of the other line
T10	else	0	Dislocation 1/3 and deletion of 2/3 based on T7
T11	T12	80	Group B benchmark data
T12	T11	80	Group B benchmark data duplicate detection
T13	T12	80	Group B benchmark data duplicate detection
T14	T11	60	Group B based on the direction D of translation of 1/10 of the other line
T15	T11	60	T14 duplicate detection
T16	else	0	Group B based on the direction U of translation of 1/10 of the other line
T17	T16	100	T16 duplicate detection
T18	T19	60	Group B similar tools, different ends
T19	T18	60	T18 duplicate detection
T20	T22	20	T18 based on the direction D of translation of 1/10 of the other line
T21	else	0	T20 duplicate detection
T22	else	0	T20 based on the direction D of translation of 1/10 of the other line
T23	T22	60	T23 duplicate detection
T24	T25	80	Group C benchmark data
T25	T24	80	Group C benchmark data duplicate detection
T26	T25	20	Group C based on the direction D of translation of 1/10 of the other line
T27	T28	60	T26 duplicate detection
T28	T27	60	T26 duplicate detection
T29	T30	60	Group C based on the direction U of translation of 1/10 of the other line
T30	T29	60	T29 duplicate detection

shown in Table 2. The calculation takes 38 s, with a matching success rate of 51% and a failure rate of 49%.

Compared with the method proposed in Ref. 8, the traceability technique proposed in this study has significant advantages in terms of operation precision and stability order. Although there are no clear differences in the speed of operation, the proposed method is more applicable to test the toolmark data in actual scene detection.

5. Conclusions

In this study, we focused on airport cable-cutting toolmark rapid-tracing matching using a new signal processing technique based on single-point laser sensing. The study was carried out on actual mark inference cable cutting tools. The boxplot was applied to perform abnormal data correction. Then, a rotation angle adaptive correction was performed to unify the matching data. Furthermore, multimatching strategies based on threshold sequence and difference

in variance were used to perform a similarity matching of toolmark features. For further validation, the experiment was conducted through an actual shear tool source experiment; the correct determination of the corresponding tools was rapidly achieved, confirming the applicability of this method to highly contaminated guided wave signals. The complexity of the algorithm presented in this study is relatively low. The algorithm can be programmed directly using the Python language, and the generated executable file can be run on mid-range computers.

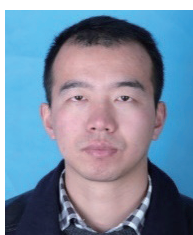
Acknowledgments

This article was supported by the Key Project of Technology Research Program funded by Yunnan Province (approved grant 2016RA042), the Innovation Fund for Technology-Based Firms in Yunnan Province (approved grants 2014SC030 and 2017EH028), the Key Science and Technology Project of Kunming City (approved grant 2015-1-S-00284), and the Ministry of Public Security Technology Research Program (approved grants 2014JSYJA020 and 2016JSYJA03).

References

- 1 S. Bunch and G. Wevers: *Sci. Justice* **53** (2013) 223.
- 2 N. Michael: *IEEE Trans. Image Process.* **3** (2006) 624.
- 3 S. Chumbley, S. Zhang, M. D. Morris, R. Spotts, and C. Macziewski: *Forensic Sci.* **62** (2017) 83.
- 4 A. B. Lock and M. D. Morris: *Technometrics* **55** (2013) 548.
- 5 C. Gambino, P. Mclaughlin, L. Kuo, F. Kammerma, P. Shenkin, P. Diaczu, N. Petraco, and J. Hamby: *Scanning* **33** (2011)272.
- 6 M. Baiker, I. Keereweer, R. Pieterman, E. Vermeij, J. V. D. Weerd, and P. Zoon: *Forensic Sci. Int.* **242** (2014) 186.
- 7 N. Pan, X. Wu, and Y. Liu: *Chinese J. Sci. Instrum.* **36** (2015) 1372.
- 8 R. Spotts, L. S.Chumbley, L. Ekstrand, S. Zhang, and J. Kreiser: *J. Forensic Sci.* **60** (2015) 303.

About the Authors



Nan Pan (from Huaiyuan, Anhui Province; Ph.D., associate professor, and master instructor) received his B.S. degree in 2008 from Anhui University of Science & Technology and his Ph.D. degree in 2012 from Kunming University of Science & Technology. He is now an associate professor and a master instructor in Faculty of Civil Aviation and Aeronautical, Kunming University of Science & Technology. His main research interests include signal processing and criminal investigation special equipment R&D. (15808867407@163.com)



Gang Li (from Shijiazhuang, Hebei Province) received his B.S. degree in 2003 from Criminal Investigation Police University of China. He is now the Deputy Director of the Institute of Forensic Science, Shijiazhuang Public Security Bureau. His main research interest includes trace evidence examination. (921205339@qq.com)



Dilin Pan (from Huaiyuan, Anhui Province; professor and master instructor) received his B.S. degree in 1982 from Anhui University of Science & Technology and his M.S. degree in 1986 from Anhui University of Science & Technology. He is now the chief engineer of Kunming SNLab Technology Co&Ltd. His main research interests include signal processing and forensic science equipment R&D. (dilinpan@camsonar.com)



Xin Shen (from Qujing, Yunnan Province; professor-level senior engineer) received his B.S. degree in 2003 and his M.S. degree in 2012 from Kunming University of Science & Technology. He is now an engineer of Metrology Center of Yunnan Power Grid Co., Ltd. His main research interests include signal processing and internet of things. (23755803@qq.com)



Xin Wang (from Kunming, Yunnan Province; senior engineer) received her B.S. degree in 1988 from Kunming University of Science & Technology. She is now an engineer of Metrology Center of Yunnan Power Grid Co., Ltd. Her main research interests include signal processing and instrumentation. (wangxin@qq.com)