

An Advanced Detection Scheme for Concealed Micro-Electronic Devices Using Classification Algorithms

Kwangyul Kim¹, Jeonghwan Lim², Songkang Kim², Junkyung Cho², and Yoan Shin¹⁺

¹ School of Electronic Engineering, Soongsil University, Seoul 156-743, Korea

² R&D Center, Eltronix Co., Seoul 152-769, Korea

Abstract. This paper proposes an advanced detection scheme for concealed electronic devices using classification algorithms in order to protect intellectual property rights. In general, if a radio wave with a specific fundamental frequency is propagated from the transmitter of a classifier to a concealed object, the second and the third harmonics will be returned as the radio wave is reflected. Using this principle, we use the k -NN (Nearest Neighbor) and the fuzzy c-means algorithm for judging diverse concealed objects. The simulation results indicate that the proposed scheme can detect electronic devices and metal devices in diverse learning data environments through efficient classification. Furthermore, it is expected that the proposed schemes can be extended to accumulate learning data of diverse devices in order to use the proposed scheme for more efficient detection of diverse concealed objects to protect intellectual property rights.

Keywords: concealed micro-electronic devices, harmonics, detection, classification, decision boundary

1. Introduction

Recently, as the knowledge-based IT society has emerged due to the innovative development of the internet and wireless mobile communications, the importance of intellectual property rights including patent rights, has been gradually magnified as competition among various research and development has become fierce. For instance in Korea, the number of applications for intellectual property rights has been rapidly increasing every year since 1986. Based on these intellectual property rights, technical competition has become fierce not only among businesses in Korea but also throughout the world. As a result, according to a report by the World Bank published in 2009, the GDP (Gross Domestic Product) of Korea was ranked fourteenth in the world, and an IMD (Institute for Management Development) report indicated that Korean scientific technology took third place in the world, while Korean technical competitiveness took fourteenth place in the world [1].

However, although the number of intellectual property rights increased as the importance of intellectual property rights as such has been magnified, the outcomes of technology development have not been efficiently managed and protected. In particular, the 2009 IMD report indicated the level of cyber security of Korea as thirty-eighth in the world, and the 2009 WEF (World Economic Forum) reported the level of protection of intellectual property rights of Korea was forty-first in the world [1]. Thus, whereas the level of research and development has been enhanced, the contents of research and development should be effectively protected; as yet, efforts to this end are quite insufficient. As a result, industrial secrets have frequently been leaked, and economic damage has been increasing. In particular, since large data can be stored recently in micro storage media such as memory sticks, the possibility that insiders may leak documents more easily by concealing micro devices has increased.

⁺ Yoan Shin Tel: +82-2-820-0632; Fax: +82-2-821-7653;
E-mail: yashin@ssu.ac.kr

As a method to tackle this issue, a proper scheme is proposed in [2] that can judge the presence of electronic devices based on the characteristics of the second and third harmonics returned from concealed objects when the transmitter has transmitted a radio wave of a specific fundamental frequency. That scheme will judge the object to be an electronic device if the magnitude of the second harmonics is larger than that of third harmonics and to be metals in the opposite cases. However, due to the effects of interference by internal and external electromagnetic waves that may occur in the process of measuring concealed objects, false alarms may occur if the objects are judged using only magnitudes of harmonics.

Hence, in order to efficiently reduce detection errors resulting from interferences by internal and external electromagnetic waves, k -NN (Nearest Neighbor) and fuzzy c -means algorithms will be used in this paper. Through computer simulations, it could be observed that the present scheme was more efficient than general schemes since it showed detection error rates of around 10%, even in electromagnetic wave interference environments. The rest of the paper is organized as follows. In Section 2, the classification algorithms considered in the proposed detection schemes are presented. The proposed schemes are introduced in Section 3, followed by the simulation results in Section 4. Finally, Section 5 concludes the paper.

2. Classification Algorithms

2.1. k -NN Algorithm

A classification algorithm used in this paper is a k -NN algorithm which is a classical data classification algorithm to assume the classes of new data based on learning data of which the classes have already been defined [3]. In particular, this algorithm is widely used in diverse machine learning areas since its usefulness and conciseness have been proved. In the case of general k -NN classification methods, Euclidean distances between predefined learning data and new data to be classified are calculated as in (1) to find k pieces of nearest learning data, and the measured new data are defined into the class that contains more learning data among the k pieces of learning data.

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_d - y_d)^2} \quad (1)$$

where $x_i (i=1, \dots, d)$ is the attribute values (i.e., features) of the pre-defined learning data and $y_i (i=1, \dots, d)$ is the attribute values of the object of which the class has not been defined.

If we have a sufficient number of samples, the majority vote of the nearest k neighbors may be better than a single nearest neighbor. The value of k should be large enough to minimize the probability of a non-Bayes decision for the unclassified point x , but at the same time it should be small enough in order that the points be close enough to x to give an accurate estimate of the posterior probabilities of the true class of x [3].

2.2. Fuzzy c -means Algorithm

Fuzzy c -means algorithm is a data-clustering technique in which a data set is grouped into n clusters with every data point in the data set belonging to every cluster to a certain degree. For example, a certain data point that lies close to the center of a cluster will have a high degree of belonging or membership to that cluster, while another data point that lies far away from the center of a cluster will have a low degree of belonging or membership to that cluster. This is based on minimization of the following objective function as in (2) [4,5].

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|\mathbf{x}_i - \mathbf{c}_j\|^2, \quad 1 \leq m < \infty \quad (2)$$

where m is any real number greater than 1, a d -dimensional vector \mathbf{x}_i is the i -th ($i=1, \dots, N$) measured data, \mathbf{c}_j is the d -dimensional center of the cluster, u_{ij} is the degree of membership of \mathbf{x}_i in the cluster j ($j=1, \dots, C$), and $\|\cdot\|$ is any norm expressing the similarity between the measured data and the center. Fuzzy

partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers \mathbf{c}_j by (3).

$$u_{ij} = \left(\sum_{k=1}^c \left(\frac{\|\mathbf{x}_i - \mathbf{c}_k\|}{\|\mathbf{x}_i - \mathbf{c}_j\|} \right)^{\frac{2}{m-1}} \right)^{-1}, \quad \mathbf{c}_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot \mathbf{x}_i}{\sum_{i=1}^N u_{ij}^m} \quad (3)$$

This iteration will stop when $\max_{ij} \{u_{ij}^{k+1} - u_{ij}^k\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m .

3. Proposed Detection Scheme

3.1. Detection of Micro-Electronic Devices

When the transmitter of a classifier has transmitted a radio wave of a specific fundamental frequency, the signals returned from concealed objects contain harmonics of which the frequency is n times the fundamental frequency when they are received. These harmonics have diverse characteristic values, depending on target objects. In general, in the case of electronic devices, the peak intensity of the second harmonics is higher than that of the third harmonics; conversely in the case of metal devices, the peak intensity of the third harmonics is higher than the peak intensity of the second harmonics [2]. Hence, in this paper, electronic devices and metal devices are distinguished by comparing the second and the third harmonics returned and received. The present section offers explanation on the environments that occur when learning data are measured in order to classify products.

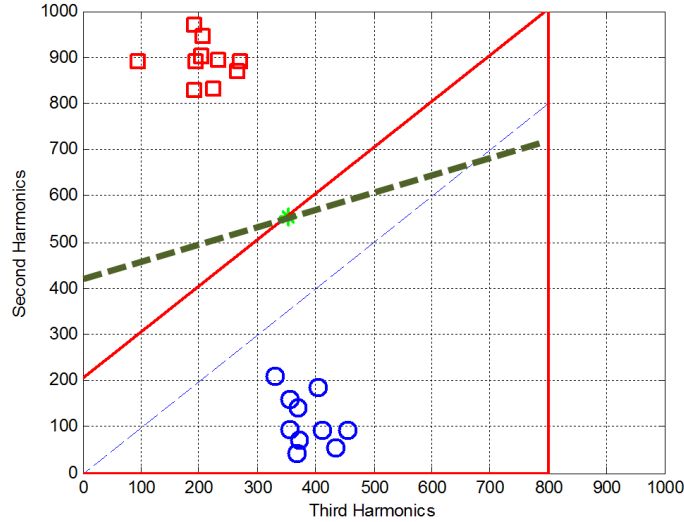


Fig. 1: Asymmetric learning data and D/B.

Figure 1 shows a situation where the learning data of the second and the third harmonics are asymmetrical. It can be intuitively seen that, if a symmetrical D/B (Decision Boundary) as shown by the thin dotted line is used when learning data are asymmetrically located, most devices may be determined to be electronic devices, and thus CEP (Classification Error Probability) may increase. In addition, in the cases where learning data are symmetrical or asymmetrical, an appropriate D/B should be set so that optimum classification can be made based on the D/B to efficiently reduce CEP resulting from interfering electromagnetic waves. Thus, considering the average values of the learning data of electronic devices and metal devices, an advanced D/B as shown by the solid line in the figure has been proposed and set. In particular, it is expected that this will enable more precise classification, even in environments with noises that may occur in the process of measuring learning data or interferences by internal and external electromagnetic waves that may occur in the process of measuring new data.

However, to provide optimum D/B's for environments with asymmetrical learning data, symmetrical D/Bs should be set based on learning data. Therefore, in this paper, a D/B scheme using fuzzy c-means algorithm was proposed to enable efficient classification, even in environments where learning data are asymmetrical.

3.2. Proposed D/B and Detection Scheme

The proposed detection scheme is composed of the following two steps. The first step is to obtain learning data and define classes based on defined D/B's using the fuzzy c-means algorithm, and the second step is to classify classes for new devices using the k -NN algorithm.

3.2.1. First Step: Fuzzy c-means Algorithm Based Decision Boundary Setting

The purpose of the first step is to make virtual symmetrical D/B's for accurate classification in environments where learning data are asymmetrical. Thus, the learning data are classified using a fuzzy c-means algorithm first. In this case, virtual learning data are used. The virtual learning data are distributed on average points of learning data values in two classes, and the fuzzy c-means algorithm classify the classes using the learning data and the virtual learning data. Hence, as shown in Fig. 1, a symmetrical D/B for the asymmetrical learning data will be formed in the center of the space between the two classes. In Fig. 1, the thin dotted line is a "Conventional D/B", the solid line is an "Advanced D/B", and finally the thick dotted line indicates the "Proposed D/B". That is, it can be seen that the "Proposed D/B" is a symmetrical D/B set considering the asymmetrical learning data.

3.2.2. Second Step: Classification and Detection

The second step is to classify the new data and check if the data indicate electronic devices. For the classification, the k -NN algorithm explained in Section 2 is used. By calculating the Euclidean distances as shown in (1) between newly measured data and individual learning data, k pieces of learning data nearest to the measurement data are found. Thereafter, the measured data are classified into the class that contains the largest number of learning data. In this case, if the class designated in advance in the first step is the same as the class determined in the second step, the detection will be considered to have been properly conducted; if not, the detection will be considered as involving errors.

4. Simulation Results

To measure the performance of the proposed detection scheme for the electronic devices, computer simulations were conducted using MATLAB. The number of learning data used was 100, consisting of 50 electronic devices and 50 metal devices; the classes of the electronic devices and metal devices were defined using all three types of D/B's explained earlier. Since the scheme is intended to reduce CEP resulting from interferences by internal and external electromagnetic waves, learning data into which interferences were intentionally included, were used. At this time, the interferences into the learning data had the characteristics of a Gaussian distribution with the mean 0 and the standard deviation of 50. The measurement data were created to have the characteristics of a Gaussian distribution against the D/B that determined two classes. The standard deviation of the created measurement data was set to 30. Since the measurement data could also be interfered by internal and external electromagnetic waves in the process of measurement, a Gaussian random variable (with standard deviation of 2) was injected into the original measurement data. The k values of the k -NN algorithm were increased by 4 pieces at a time in a range of 3~51 during the measurement; the measurement was conducted 100,000 times.

Figure 2 shows CEPs according to k in environments where learning data are asymmetrical. It can be observed that, in the case of the "Conventional D/B", CEP's of around 50% were obtained since efficient classification is possible only in symmetrical environments. However, in the cases of all schemes other than the "Conventional D/B", CEP's decreased as k increased. CEP's decreased by around 2% in the case of the "Advanced D/B" and around 5% in the case of the "Proposed D/B". When the schemes were compared for $k = 3$, CEP's decreased further by around 6% in the case of the "Proposed D/B" than in the case of the "Advanced D/B"; in cases where $k = 51$, CEP's decreased further by around 8%. In particular, in the case of the "Proposed D/B", the excellence of the proposed scheme could be identified. The reason is that

environments where learning data were asymmetrical were changed into virtual symmetrical environments to obtain results similar to the results from the k -NN algorithm, and that relative margins were given to spaces between the learning data to enable accurate classification, even in environments with the interferences by internal and external electromagnetic waves.

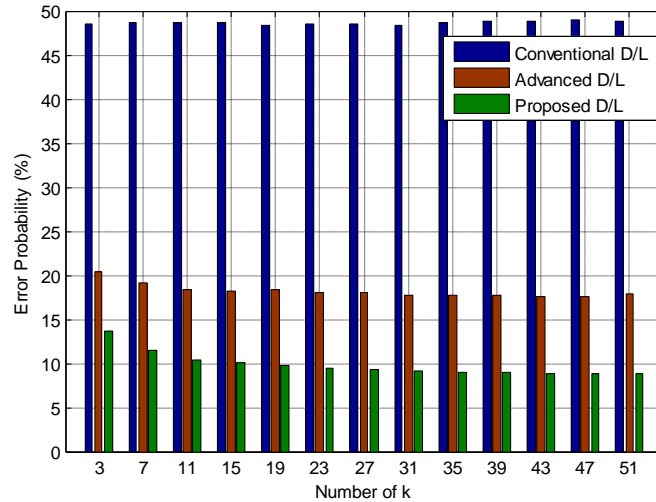


Fig. 2: CEP performance comparison of the proposed D/B with the conventional D/B and advanced D/B in the asymmetric environments.

5. Conclusion

This paper proposed an advanced detection scheme to improve the classification performance of micro electronic device detectors being developed in order to protect the intellectual property rights. To measure the performance of the proposed scheme, diverse D/B's that could define learning data classes in advance were set; and based on these D/B's concealed objects were classified using the k -NN algorithm. In particular, cases where the second and the third harmonics of learning data were considered and the effects of interferences by internal and external electromagnetic waves that might occur in the process of detecting concealed objects were also considered. From the computer simulations, it was observed that the proposed D/B scheme provided CEP reductions of around 9% in asymmetrical environment. Hence, it could be seen that the proposed scheme could not only detect concealed microelectronic devices but also show efficient classification performance, even under the effects of interference by internal and external electromagnetic waves.

6. Acknowledgements

This work was supported by the Industrial Convergence Source Technology Development Program (No. 10039192) funded by the MKE, Korea.

7. References

- [1] Korea Industrial Technology Association. *A Study on the Current Status of Overseas Industrial Security Management and Case Studies of American Companies*. Small and Medium Business Administration of Republic of Korea, 2009.
- [2] B. R. Barsumian, T. H. Jones, and C. A. Liter. Surveillance device detection utilizing nonlinear junction detection and reflectometry. *US Patent #7,212,008*, 2007.
- [3] T. M. Cover, and P. E. Hart. Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory*. 1967. **IT-13** (1): 21-27.
- [4] F. Palumbo, C. N. Lauro, and M. J. Greenacre. *Data Analysis and Classification*. Springer, 2010.
- [5] S. Theodoridis, and K. Koutroumbas. *An Introduction to Pattern Recognition: A MATLAB Approach*. Academic Press, 2010.