Progress In Electromagnetics Research M, Vol. 30, 211–224, 2013

A SUPPORT VECTOR MACHINE FOR IDENTIFICATION OF MONITORS BASED ON THEIR UNINTENDED ELECTROMAGNETIC EMANATION

Fan Mo, Yinghua Lu*, Jinling Zhang, Qiang Cui, and Sihai Qiu

School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract—Electrical equipments usually radiate unintended emission which carries characteristic information when running, such as emanation from computers monitors, keyboards and other components, this emanation can be possibly used to reconstruct the source information. Most of the experiments related to this area are carried out inside a semi-anechoic chamber, and measurement out of it may not be considered to be optimal, because the data captured are usually not sufficient. Yet in this study, we take LCD monitors as typical examples and find that characteristics significantly differ between products, parameters such as the magnitude and spectrum were measured under normal environment. We take the PCB traces as antennas and acquire the raw signal directly near the antenna and extract the parameters to use as input to support vector machine (SVM) which was trained to identify the emanating source (LCD monitors). In this study, the method was tested using the emission captured from one Samsung (SyncMaster E1920) and two LG (L1753s) monitors, and a laptop (ACER Aspire 5542). The SVM was able to classify the source of signals with 98.9510% accuracy while using emission that captured from the running monitors.

1. INTRODUCTION

Emanation from electrical equipments can be used for electromagnetic eavesdropping. Taking computer monitor as example, in 1985 the first eavesdropping of the cathode-ray tube (CRT) monitors was demonstrated by van Eck. Experiments found that in proper

Received 24 December 2012, Accepted 18 April 2013, Scheduled 27 April 2013

* Corresponding author: Yinghua Lu (yhlu@bupt.edu.cn).

environment, the text and image on the display can be reconstructed. Modern monitors such as LCD, LED screens, which seem to have less electromagnetic emanation than CRT, are still able to be eavesdropped nearby, especially nowadays the transmission speed of modern digital products has risen to Gbit/s, which in a way make the eavesdropping much easier and the eavesdropping equipment hard to discover [1, 2].

For different monitor products as well as the monitors of the same product, due to their different components structure and driving circuit frequency, the reconstruction methods are different. In the latter case, the difference turns out to be much more subtle, and we assume that method used to reconstruct is the same. Literatures show that when the monitor parameters are known by detail the reconstruction of the display becomes available [3–5]. Yet little research mentions how to respond when the detail of the target is not confirmed. Therefore, in this study we need to acquire the emanation, with which we can not only distinguish monitors of different products, but also share the same characteristic with different monitors of the same products. In order to successfully extract these characteristics, we need to properly collect the feature of each monitor from the spectrum, and use algorithms to detect, identify, and locate the monitors automatically.

Figure 1 shows the emanation spectra of two different products of monitors (Samsung SyncMaster, ACER) compared with the ambient noise measured in the Electromagnetic Compatibility (EMC) Laboratory at Beijing University of Posts and Telecommunications (BUPT). This emanation was measured using near-field probes connected to a spectrum analyze, in which resolution bandwidth is

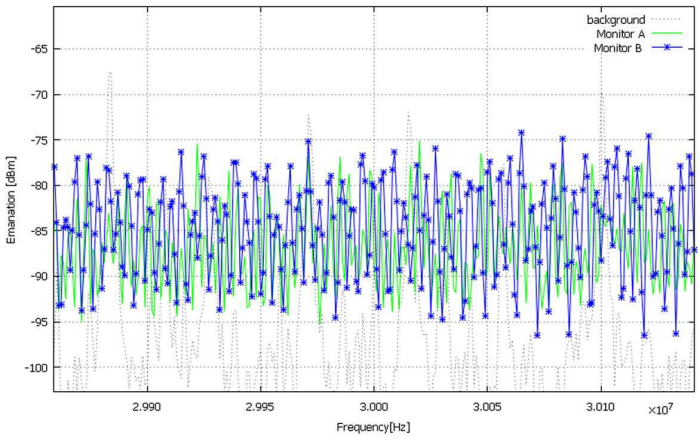

Figure 1. Emanations spectra of two different monitors.

set to 1 kHz. The figure shows that the characteristics of different monitors vary significantly from one to another, and the differences can be used to train the learning machine to automatically identify the monitors based on the characteristic of the emanation.

To identify the emission source, pattern identification technique can be widely implemented in this area. Some typical nonlinear algorithms, such as artificial neural networks (ANNs), support vector machine (SVM), have been commonly used in target recognition. Both of these learning machines are powerful, efficient, and robust in solving these problems. They are able to be trained from learning set and to generalize the target characterize accurately [6–13].

On the other hand, there is a difference between these two techniques, which leaves an option to the researchers when deciding which algorithm to apply, e.g., the ANN algorithm is based on empirical risk minimization (ERM) principle, while the SVM use the structural risk minimization (SRM). Which means, ANN is highly depended on the training data set, in this way it lead to two problems, a) the quality of the training data set directly affects the accuracy of the machine. To acquire high quality training set, abundant data set is needed; therefore the learning time and complexity of the calculation are inevitably increased; b) the accuracy of the machine also depends on the experience and priori assumption of the learning machine designer. In this way, the results of the machine are different from person to person. However, SVM, which is strictly based on the statistic and mathematic theory, does not require the designers' experience and priori assumption. Moreover, SVM is based on small sample statistical learning theory, thus compared with ANN, it is more efficient in classifying the target correctly when the training data set is limited. In addition, due to this computational advantage, it has better performance in calculation speed and flexibility. While dealing with the non-linear case, SVM offers a more accurate solution to the target identification problems, in which linear algorithm can not [14, 15].

2. METHOD

2.1. Radiation Model

According to electromagnetic theory, the electromagnetic wave, which is caused by moving charges or changing current, will emit and spread to the space. Whereas in the digital equipments, the signals consist largely of periodic trapezoidal waveform and impulse signal, of which the change of rising edge and falling edge can be taken into consideration in antenna models. For the length of the conductor L,

if it is far less than the length of the microwave in the vacuum, it can be taken as a standing wave dipole model.

In the experimental environment, the difference of the monitors' radiation is mainly caused by its own circuit structure, driving clock, and the graphic cards. In 2003, Kuhn conducted an experiment with different graphics adapters and measured the spectrum of it showed as Fig. 2 [16].

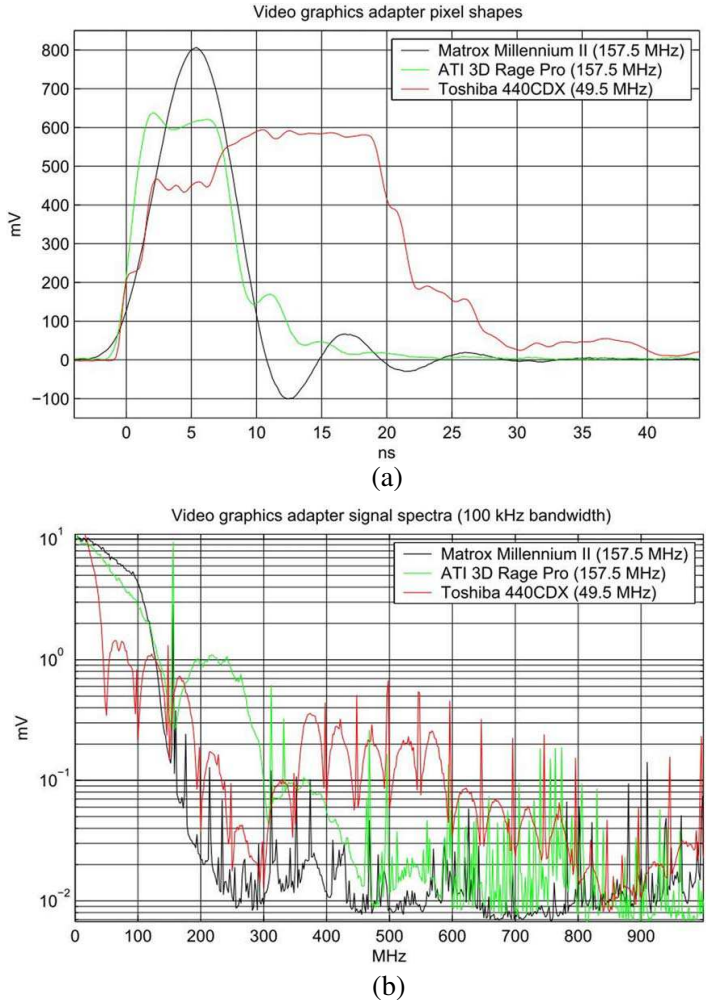

Figure 2. (a) Voltage curve of a single full-intensity pixel for two graphic cards and a laptop. (b) Frequency spectrum measured for a random-bit test image with a spectrum analyzer.

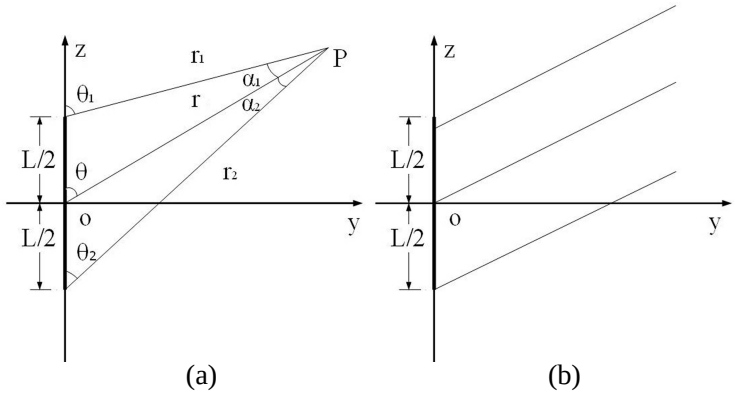

Figure 3. (a) Geometry of a linear antenna in near zone. (b) Geometry of a linear antenna in the far zone.

From the figures we can see that compared with different graphic adapters, the power is mainly at the frequency of 600 MHz or even lower than this frequency, and the length of the microwave in free space λ is 0.5 m or longer than this, which means that the length of the conductors in the monitor circuits is suitable for the standing wave dipole model. Thus in this model, the current density on the conductor can be considered as constant at the same moment:

$$J(z,t) = \begin{cases} \text{constant} & -\dfrac{L}{2} \le z \le \dfrac{L}{2} \\ 0 & \text{others} \end{cases} \tag{1}$$

The radiation of electric equipments includes static electric field, electric and magnetic fields. In the near zone, the dominant fields are the static electric field and magnetic field, while in the far zone is the electric field. The far zone is defined as a zone of which the distance r from the source is far greater than $\frac{\lambda}{2\pi}$. At the distance near $\frac{\lambda}{2\pi}$, the zone is a composite zone, and the zone, of which the distance is far less than $\frac{\lambda}{2\pi}$, is defined as near zone. In the near zone and far zone, the electric and magnetic fields are shown as Fig. 3 [17]. And the electric and magnetic fields are equal to:

$$\begin{aligned} E_\theta \;=\; & \frac{\eta_0}{2\pi\rho}\left[J\left(t-\frac{r}{c}\right)+J\left(t-\frac{L}{c}-\frac{r}{c}\right)\right. \\ & \left.-J\left(t-\frac{L}{2c}-\frac{r_1}{c}\right)\cos\alpha_1-J\left(t-\frac{L}{2c}-\frac{r_2}{c}\right)\cos\alpha_2\right] \end{aligned} \tag{2}$$

$$E_r = \frac{\eta_0}{2\pi\rho}\left[J\left(t-\frac{L}{2c}-\frac{r_1}{c}\right)\sin\alpha_1 - J\left(t-\frac{L}{2c}-\frac{r_2}{c}\right)\sin\alpha_2\right] \quad (3)$$

$$H_\varphi = \frac{1}{2\pi\rho}\left[J\left(t-\frac{r}{c}\right)+J\left(t-\frac{L}{c}-\frac{r}{c}\right) - J\left(t-\frac{L}{2c}-\frac{r_1}{c}\right)-J\left(t-\frac{L}{2c}-\frac{r_2}{c}\right)\right] \quad (4)$$

At the distance where $r \to \infty$, the electric and magnetic field of the standing wave dipole mode are represented as equations below:

$$E_\theta = \frac{\eta_0}{2\pi\rho}\left[J\left(t-\frac{r}{c}\right)+J\left(t-\frac{L}{c}-\frac{r}{c}\right)-J\left(t-\frac{L\left(1-\cos\theta\right)}{2c}-\frac{r}{c}\right) -J\left(t-\frac{L\left(1+\cos\theta\right)}{2c}-\frac{r}{c}\right)\right] \quad (5)$$

$$E_r = 0 \quad (6)$$

$$H_\varphi = \frac{1}{2\pi\rho}\left[J\left(t-\frac{r}{c}\right)+J\left(t-\frac{L}{c}-\frac{r}{c}\right)-J\left(t-\frac{L\left(1-\cos\theta\right)}{2c}-\frac{r}{c}\right) -J\left(t-\frac{L\left(1+\cos\theta\right)}{2c}-\frac{r}{c}\right)\right] \quad (7)$$

As mentioned above, the waveform of signal in digital products are mainly trapezoidal waveforms, of which the difference is shown as Fig. 4. According to the standing wave dipole mode equations presented above, while in the near zone, the radiated trapezoidal signals with different ΔT_i, which is caused by the different durations of the rising edge and falling edge, share the same ΔT_0. However, in the far zone, they are different both in magnitude and time span, and this character can be used to distinguish the different signals.

In order to capture the signals efficiently, we also need to figure out how the coupling models work, which includes the coupling between the microwave in the free space and the receiver antenna, the coupling between the microwave and the transmission lines, and the coupling among the cables.

(i) Inductive coupling
When the transmitting signals are in one of the parallel lines, due to the coupling between this pair of lines, in the line of another, the signal induced from the source line will cause information leakage problems. The relation of voltages in these two lines can be probably described as:

$$V_i = \frac{j\omega Ml}{2}V_s \quad (8)$$

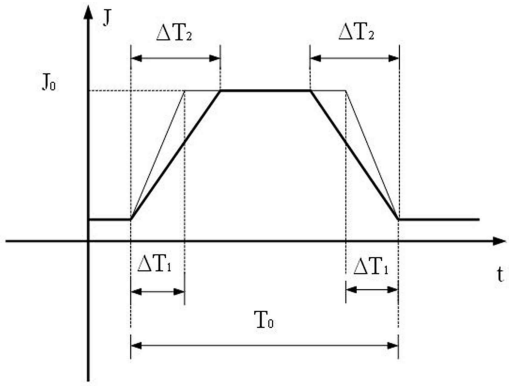

Figure 4. Trapezoidal waveform signal in digital products.

where V_i is the inductive voltage from the source voltage V_s, M the coefficient of mutual induction, l the length of the coupling lines, and ω the frequency of the transmission signal. V_i is proportional to M, l and ω. In most cases, since the signal frequency is confirmed by default, M and l are the only optional factors to deal with, e.g., increasing the distance between two lines can efficiently restrain the coupling.

(ii) Conductive coupling
At high frequency, the main coupling of the parallel lines becomes the conductive coupling. In high speed circuit, if the traces on the PCB layout are not properly designed, the signals on it will be coupled to the other traces. Therefore, the shielding of the high-frequency traces is necessary. While in the conductive coupling, the relation of voltages between the circuits is approximately described as:

$$V_c = j\omega C R_l V_s \tag{9}$$

where V_c is the conductive voltage and R_l the load resistance. The coupled conductance C equals to:

$$C = \frac{\pi\varepsilon_0}{\cosh^{-1}(\frac{d}{D})}$$

The distance between two traces and the diameter are respectively represented by d and D.

(iii) Coupling due to the shared impedance
While two equipments share the common return, e.g., grounding, disturbance or coupling will occur. The most common situation is grounding coupling and the power coupling. Information leakage

through these shared routes can also be captured and cause security problems. The relation of the voltages in this model is approximately described as:

$$V_r = \xi V_s \tag{10}$$

and ξ equals to:

$$\xi = \frac{R_r}{r_r + Z_r + R_r} \frac{R_s}{r_s + Z_s + R_s}$$

where R_r, R_s are the load resistances of the coupled received circuit and source circuits respectively. Also, r_r, r_s are the internal resistance, and Z_r, Z_s the output impedance.

2.2. Measurement and Identification Process

As mentioned above, from Fig. 2, we can see that the main power of radiation is concentrated at the bandwidth of $0 \sim 600$ MHz. To receive signals at different frequencies, different types of antennas are chosen. For example, when the narrow-band signal interested is well estimated, we can use simple dipole antenna to receive it if its bandwidth is not more than the $\frac{1}{10}$ of the center frequency of the antenna. In addition, Yagi antennas also have excellent performance in receiving the narrow-band signal, especially at VHF and UHF when the frequency center of the eavesdropping target is precisely known.

While referring to broad-band antenna, from low frequency to ultra-high frequency, there are common antennas used for compromising emanation eavesdropping, e.g., active monopole antenna works in the bandwidth at 100 Hz $\sim$ 30 MHz; bi-conical antenna at 30 MHz $\sim$ 300 MHz; log-periodic antenna at 200 MHz $\sim$ 1000 MHz, and discone antenna at 200 MHz $\sim$ 1300 MHz.

In this research, the clocks on the monitors work at the frequency varied from low frequency to about 500 MHz, which means the length of microwave radiated $\lambda \geq 0.6$ m, thus we take the traces on the PCB layout of the display as wire antennas. The model we use has been discussed in Section 2.2.1. We focus on the VHF and use near field probe to receive the signal directly near the antenna, and the distance between them is about 5 cm $\sim$ 10 cm.

We choose a clean bandwidth at 30 MHz with about -100 dBm average of the measurement ambient noise. On this bandwidth there is only a stable unknown signal with a resonant frequency of 44 kHz; therefore, it meets the needs of acquiring the characteristic signal of the target. The resolution of each monitor has been taken into consideration, and we use the same configuration to find the distinction of different monitors. Fig. 1 shows the spectrum at 30 MHz with

400 kHz bandwidth and 1 kHz resolution bandwidth. We also capture a spectrum with 80 kHz bandwidth at the same frequency. Accordingly, the resolution bandwidth is set to 200 Hz, and all 336 measurements are recorded. Spectrum of a Samsung monitor and a laptop is shown in Fig. 5.

The process of the identification using SVM is shown as Fig. 6. There are several types of basic kernel function in SVM, and we choose RBF kernel for two reasons. First of all, the RBF nonlinearly maps sample into a higher dimensional space, thus it is capable of handling

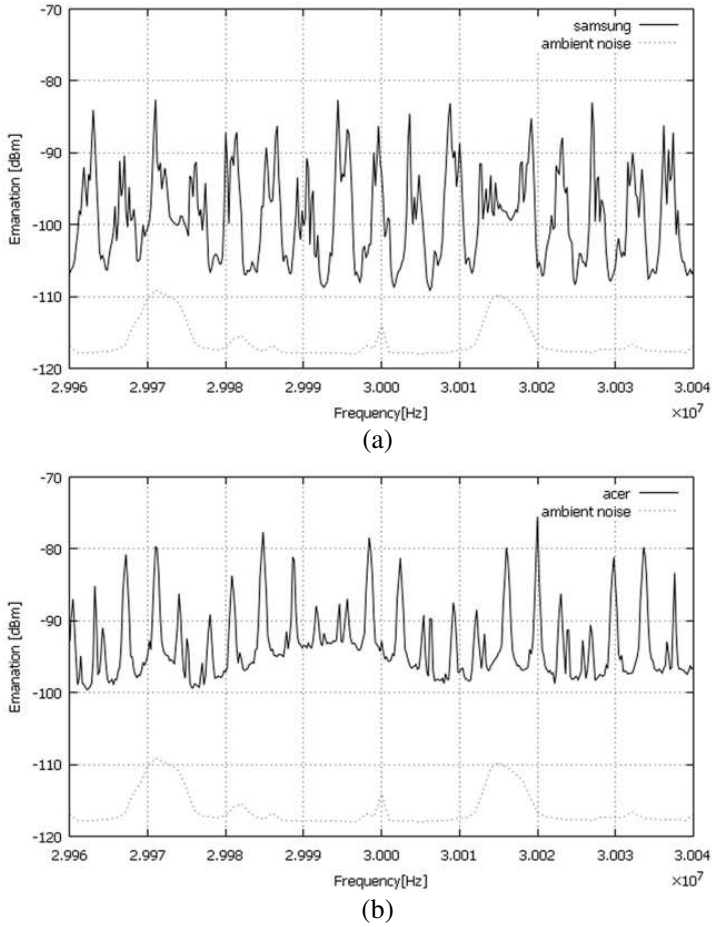

Figure 5. (a) Spectrum of a Samsung monitor with the ambient noise. (b) Spectrum of a laptop. Both of the curves are the average trace while capturing using the spectrum analyzer.

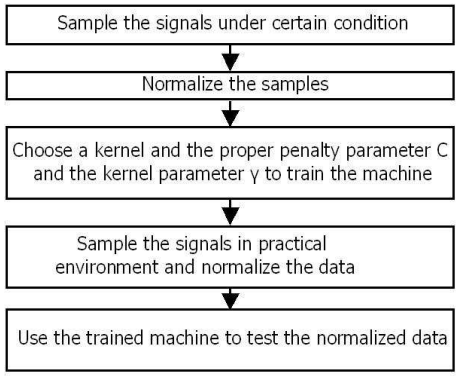

Figure 6. The process of identification using the SVM approach.

the cases of which the relation between cases labels is nonlinear. Furthermore, the RBF kernel function has fewer numerical difficulties, which makes it suitable for this case.

3. IDENTIFICATION RESULTS AND DISCUSSION

According to the measurements, for those monitors of different products, the spectrums differ significantly from one to another. Fig. 7 shows the difference of each monitor emanations in magnitude.

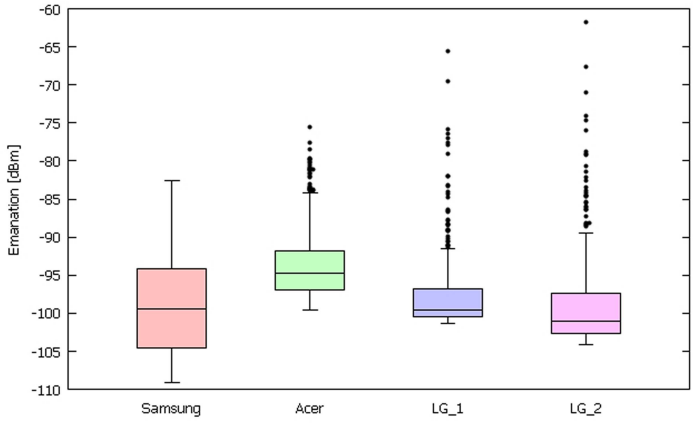

Figure 7. Boxplot of emanation of different monitors captured at 30 MHz with 80 kHz bandwidth.

In Fig. 5, we can see that the curves on the spectrum of different monitors are different from each other in shape. Besides, if these two spectrums are divided equally into frequency bands, the maximum and minimum magnitudes over these bands are all different by about $-5 \sim -10$ dBm respectively. Additionally, the average magnitude of each frequency band can also be used as the training characteristics too. Fig. 7 shows the difference of using boxplot. For monitors with different products, e.g., Samsung, Acer (laptop) and LG, we can distinguish them from the range and distribution of their values; monitors with the same products, shown as LG_1 and LG_2, are similar to each other both in magnitude and distribution.

3.1. Monitors of Different Products

To train the SVM machine to identify the source emanation, we choose the spectrum of different emissions with the same center frequency, bandwidth, resolution bandwidth and sample rate. Also we extract 401 examples in each spectrum to confirm that each training sample has the same scale. On the first step, we choose monitors of three different types of products which are Samsung, Acer, LG (mentioned as LG_1) to be the identification target, and we also use a different LG monitor of the same product (mentioned as LG_2) to perform the unknown and irrelevant type of signals. Secondly, when the monitors run stably, we have sampled 50 groups of random emanation from three monitors (Samsung, Acer, LG_1) and 10 groups of LG_2 to train the SVM machine. Thirdly, we use both the groups of totally different data, e.g., the capture of the average trace using the spectrum analyzer, and some random groups of trained data to be the identification target of the machine. The training results are shown in Table 1.

Table 1. Accuracy (correct numbers: whole samples) to identify the monitors of the different products.

Samsung	Acer	LG_1	Others	Hit rate
37 : 37	7 : 7	20 : 20	0 : 1	98.4615% (64 : 65)
25 : 25	24 : 24	49 : 49	3 : 4	99.0196% (101 : 102)
42 : 42	13 : 13	3 : 3	7 : 8	98.4849% (65 : 66)
21 : 21	18 : 18	11 : 11	3 : 3	100% (53 : 53)

In Table 1 the numbers of the sample used for detection are generated by random functions. While training the SVM machine, we use kernel RBF, and to increase the accuracy, we choose γ with

0.0078125 and the penalty parameter C with 2 after cross validation, and get an average accuracy of 98.9510% after 286 machine validation.

There are 3 times of fail validation, and to be specific, as we mentioned above. The unknown and irrelevant signals use the sample data of LG_2. All of these fail validations are misjudged LG_2 as LG_1, which means that the emanation from the monitors differs from product to product while sharing the same features of the same product.

3.2. Monitors of the Same Products

On the other hand, when training the machine, to decrease the fail validation rate we consider including the monitors sample of the same product, and similarly we use random function to generate the numbers of groups used for training. The results are shown as Table 2.

Table 2. Accuracy (training numbers: validating samples) to identify the monitors of the same products.

C, γ	Samsung	Acer	LG_1	LG_2	Hit rate
8, 0.0078125	31 : 19	36 : 14	13 : 37	2 : 8	96.1538% (75 : 78)
128, 0.0001221	13 : 37	27 : 23	36 : 14	5 : 5	100% (79 : 79)
2, 0.0078125	25 : 25	41 : 9	33 : 17	7 : 3	100% (54 : 54)
2, 0.0078125	18 : 32	18 : 32	43 : 7	6 : 4	100% (75 : 75)

Table 2 shows the results using random quantities groups to train and validate the samples and get an average accuracy of 98.9511%. As expected, the fail validation also comes from the group of LG_2 which is misjudged as LG_1. Compared with 1, the hit rate is increased.

4. CONCLUSION

A study to analyze the emanation model monitors and the method to detect and identify the display electromagnetic emissions are presented. All the emanation samples are captured from the normal office environment, and SVM is used for the characteristic identification. Parameters such as the magnitude, average and the deviation are extracted and used to support the SVM machine. In all the training data and validating samples, the ambient noise is considered. According to the experiments results, it is found that emanations from monitors of different products differ significantly while those from monitors of the same product share similar characteristics. These

characteristics can be distinguished by SVM machine with an accuracy of 98.9510%.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 61072136, 61171051 and Doctorate Foundation of the Ministry of Education of China under Grant 200700130046.

REFERENCES

1. Petitcolas, F. A. P., "Information hiding-a survey," *Proceedings of the IEEE*, Vol. 87, No. 7, 1062–1078, 1999.
2. Kuhn, M. G., "Electromagnetic eavesdropping risks of flat-panel displays," *Privacy Enhancing Technologies*, Vol. 3424, 88–107, 2005.
3. Kuhn, M. G., "Compromising emanations of LCD TV sets," *IEEE International Symposium on Electromagnetic Compatibility*, 931–936, 2011.
4. Kuhn, M. G., "Optical time-domain eavesdropping risks of CRT displays," *IEEE Symposium on Security and Privacy*, 3–18, 2002.
5. Kuhn, M. G., "Soft tempest: Hidden data transmission using electromagnetic emanations," *Lecture Notes in Computer Science*, Vol. 1525, 124–124, 1998.
6. Dudczyk, J., "Applying the radiated emission to the specific emitter identification," *15th International Conference on Microwaves, Radar and Wireless Communications*, Vol. 2, 431–434, 2004.
7. D'Amore, M., "A neural network approach for identification of EM field sources: Analysis of PCB configurations," *IEEE International Symposium on Electromagnetic Compatibility*, Vol. 2, 664–669, 1998.
8. Aunchaleevarapan, K., "Classification of PCB configurations from radiated EMI by using neural network," *Asia-Pacific Conference on Environmental Electromagnetics*, 105–110, 2000.
9. Angiulli, G., "Support vector regression machines to evaluate resonant frequency of elliptic substrate integrate waveguide resonators," *Progress In Electromagnetics Research*, Vol. 83, 107–118, 2008.
10. Tokan, N. T., "Knowledge-based support vector synthesis of the microstrip lines," *Progress In Electromagnetics Research*, Vol. 92, 65–77, 2009.

11. Tokan, N. T., "Support vector characterisation of the microstrip antennas based on measurements," *Progress In Electromagnetics Research B*, Vol. 5, 49–61, 2008.
12. Wang, F. F., "The support vector machine for dielectric target detection through a wall," *Progress In Electromagnetics Research Letters*, Vol. 23, 119–128, 2011.
13. Xu, Y., "A support vector regression based nonlinear modeling method for SiC MESFET," *Progress In Electromagnetics Research Letters*, Vol. 2, 103–114, 2008.
14. Vapnik, V., *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
15. Chang, C.-C., "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 27, 1–27, 2011.
16. Kuhn, M. G., *Compromising Emanations: Eavesdropping Risks of Computer Displays* , Citeseer, Cambridge, 2003.
17. Podosenov, S. A., "Transient radiation of traveling waves by wire antennas," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 37, No. 3, 367–383, 1995.