



HAL
open science

Enhanced THz tags authentication using multivariate statistical analysis

S. Salhi, Florent Bonnefoy, Stéphane Girard, Maxime Bernier, Nicolas Barbot, Romain Siragusa, Etienne Perret, Frédéric Garet

► **To cite this version:**

S. Salhi, Florent Bonnefoy, Stéphane Girard, Maxime Bernier, Nicolas Barbot, et al.. Enhanced THz tags authentication using multivariate statistical analysis. IRMMW-THz 2019 - 44th International Conference on Infrared, Millimeter, and Terahertz Waves, Sep 2019, Paris, France. pp.1-2. hal-02282841

HAL Id: hal-02282841

<https://hal.science/hal-02282841v1>

Submitted on 10 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enhanced THz tags authentication using multivariate statistical analysis

S. Salhi^{1,2}, F. Bonnefoy¹, S. Girard², M. Bernier¹, N. Barbot³, R. Siragusa³, E. Perret³ and F. Garet¹

¹Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, Grenoble INP, IMEP-LAHC, 73376 Le Bourget du Lac, France.

²Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

³Univ. Grenoble Alpes, Grenoble INP, LCIS, 26000 Valence, France.

Abstract—In this paper, we report on the unitary authentication of identically realized diffraction grating-based tags structures in the THz domain by using multivariate statistical analysis. We proceed to a dimension reduction with Principal Component Analysis (PCA) as a preprocessing step before a Gaussian classification and we evaluate the error rates. We then classify the tags using a Linear Discriminant Analysis (LDA). We demonstrate that PCA gives average error rates lower than 0.5% whereas LDA is able to classify the tags with error rates lower than 6.10^{-5} considering its 3 first axes.

I. INTRODUCTION

Principal Component Analysis (PCA) is mainly used for the classification of materials and authentication in various applications domains [1],[2] and has recently attracted research in the THz domain [3],[4]. This method is also commonly used to reduce the dimension of a statistical set of samples [5]. However, more efficient multivariate statistical methods have been developed specifically for classification like the Linear Discriminant Analysis (LDA) [5]. In this study, we propose to use both PCA and LDA to address the authentication of THz tags with the aim to fight counterfeiting of manufactured goods. First, PCA is used to both reduce the dimensions of the statistical measurements set and classify the different tag signatures. Secondly, LDA is used to increase the discriminative rate of these tags. The THz tag structure used in this study is schemed in Fig. 1. A 1D rectangular diffraction grating is engraved on the surface of a plastic card. The “substrate” layer acts as a dielectric waveguide whose spectral response depends on the geometrical parameters of the grating engraved on its top. In the present case these parameters are: period $\Gamma = 800 \mu\text{m}$, groove’s depth $p = 280 \mu\text{m}$, card thickness $e_{card} = 775 \mu\text{m}$, card refractive index $n_{card} = 1.73$ and absorption coefficient $\alpha = 17 \text{ cm}^{-1}$ @ 1 THz. We use the presence of guided modes appearing as narrow lines in the transmitted spectra for example, more precisely their frequency positions and width, as a potential unique signature for the tag [7],[8]. In order to prove that such a signature can be used to unitary authenticate a tag, 44 tags were realized with the same pattern and protocol; these tags can be supposed as theoretically identical. However, due to random uncertainties during the

fabrication process, each signature presents tiny variations that we take advantage using PCA and LDA.

II. RESULTS

Each tag was repeatedly measured in transmission one hundred times using classical THz-TDS system in the frequency range 200 GHz – 800 GHz (spectral range of the maximum dynamics), with a frequency resolution $\delta f = 8.6$ GHz and under an incident angle $\theta = 10^\circ$.

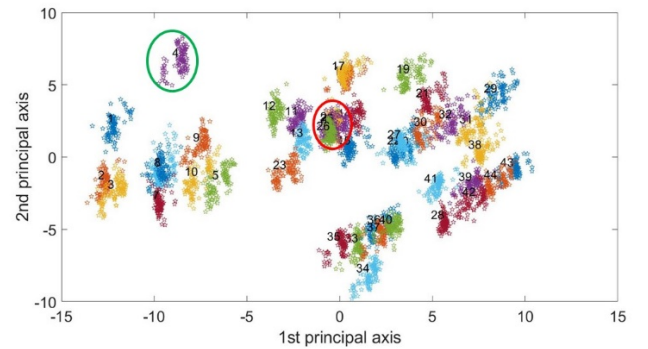


Fig. 2: Projections on the principal component space of the hundred measurements performed on each tag. For sake of clarity, each set of measurements has a different colour to its closest neighbours letting appear clusters, which are referenced by the tag number. The explained variances are 68% and 22% for the 1st and 2nd principal components, respectively.

Here, we thus considered the 4400 measurements of the twin tags, each containing 70 frequency points, as a statistical dataset thereby composed of $3.08 \cdot 10^5$ samples. We then first calculate the covariance matrix of this dataset [5]. Afterwards we diagonalize this matrix to obtain the associated eigenvalues and eigenvectors. From these latter, it is possible to get the total explained variance and the privileged directions of maximum variance, which are known as the principal components axes in PCA. The hundred extracted spectral responses of each tag were then projected on the two first principal components of PCA, corresponding to the maximal variance (Fig. 2). Each set of hundred measurements leads to a cluster whose probability density can be fitted by a bivariate normal distribution. Note that the cluster size is directly related to the measurement reproducibility. We proceed to a dimension reduction with PCA as a pre-processing step before a Gaussian classification and we evaluate the error rates. For that, a Gaussian classifier is then implemented using this bidimensional model [9]. The error probability $\epsilon_{FN}^{k,t}$ to wrongly authenticate the tag k as the tag t (False-Negative case, FN), is directly linked to the clusters size and proximity. $\epsilon_{FN}^{k,t}$ is numerically calculated from the N measurements of the k^{th} tag using:

$$\epsilon_{FN}^{k,t} = \frac{1}{N} \sum_{i=1}^N \{f_k(x_i) < f_t(x_i)\}, \quad x_i \sim f_k(x) \quad (1)$$

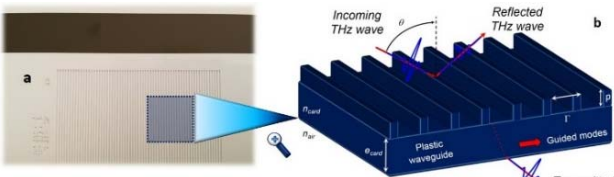


Fig. 1: a, Picture of the proposed tag and b, scheme of the structure.

With $f_{k,t}(x_i)$: the bivariate normal distribution associated to the tag k and tag t . The probability E_k of confusing the k^{th} tag with any others is therefore given by:

$$E_k = \frac{1}{M} \sum_{t=1}^M \varepsilon_{FN}^{k,t}, \quad t \neq k \quad (2)$$

where $M=44$ and $N=100$ in the present case. As expected (cf. Fig. 2), isolated clusters present the lowest error rates (e.g. $E_4 < 0.02\%$, green ellipse), while clusters packed each other are more hardly authenticated due to their close vicinity (e.g. $E_{25} < 2.7\%$, red ellipse). We found a global error rate, calculated by averaging the error rates E_k of 0.5%. This means that PCA proposed in this study can unitary authenticate 1 tag randomly chosen among 44 others, with a success score of 99.5%. In other hand, we show that the 5 first principal axes can explain 99.5% of the total variance. In such a case, the dimensions of the statistical problem could be then reduced by about a factor 15, with an information loss lower than 0.5%.

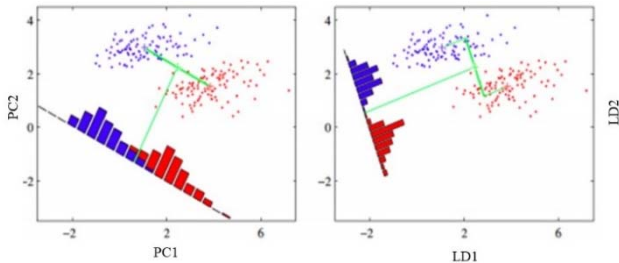


Fig. 3: Projections of two different classes (blue and red dots) in the principal component space from PCA (1st and 2nd axis, left) and on the two first axes from LDA (1st and 2nd axis, right).

We now use LDA [10] to try to better separate all the different classes. LDA permits to find the optimal base to project the clusters (blue and red scatter plot in Fig. 3). Let notice that using LDA, the obtained classes, corresponding to such projections can be much better separated than using PCA (see blue and red Gaussian curves). On the other hand, PCA remains necessary to reduce the dimension of the statistical sample before the classification process but not for efficient classification. In Fig. 4, we also plot the rate of correct classification versus the number of axis we take into account for classification. Let notice that this rate can reach 99.994% using only the 3 first axis from LDA. To evaluate the robustness of the classification, we considered a training database constituted of 70% from the original 4400 measurements while the other 30% will constitute the test database. We then calculate the correct classification rate of each sample in the test database. The whole process is randomly repeated 30 times to test the robustness of the classification process. Thereby we also showed that in these conditions, only 6% of the 4400 measurements, with at least 1 measurement on each tag, is sufficient to obtain the previously mentioned classification rate. Let notice that this result can be different if changing the original database measurements *i.e.* the set of tags and/or THz reader.

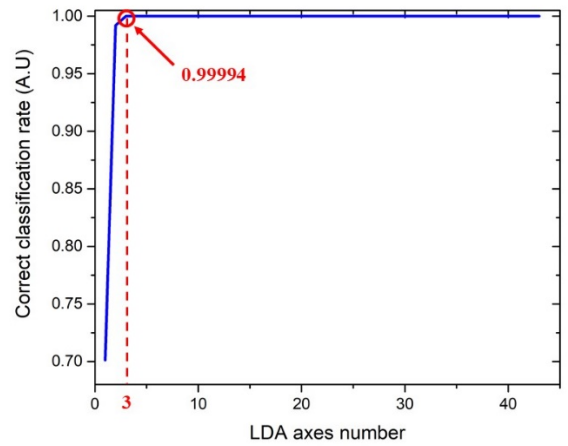


Fig. 4: Correct classification rate versus the number of LDA axes considered.

III. CONCLUSION

In this study, we demonstrate the ability to use dielectric planar waveguides associated with a diffraction grating as a tag for unitary authentication in the THz domain. More precisely we take advantage of the very small geometrical differences ($\lambda/10 - \lambda/100$) intrinsically and randomly induced by the fabrication process. To analyze the tags signatures we combined PCA and LDA to respectively reduce the dimension of the statistical sample and to efficiently classify them. A correct classification rate of 99.994% has been obtained.

ACKNOWLEDGMENT

This work is supported by the ANR AUSTRALE project, grant ANR-18-CE39-0002 of the French Agence Nationale de la Recherche.

REFERENCES

- [1] C. Güler, G. D. Thyne, J. E. McCray and K. A. Turner, « Evaluation of graphical and multivariate statistical methods for classification of water chemistry data », *Hydrogeology Journal*, vol. 10, n° 4, p. 455–474, août 2002.
- [2] M. Behera and V. K. Govindan, « Palm print Authentication Using PCA Technique », *International Journal of Computer Science and Information Technologies*, vol. 5 (3), p. 3638-3640, 2014.
- [3] S. Sommer *et al.*, « Terahertz Spectroscopy of Aged Epoxy Resin Adhesive », *9th THz Days Conference, Dunkirk, France*, 2017.
- [4] Y. Xie et P. Sun, « Terahertz data combined with principal component analysis applied for visual classification of materials », *Optical and Quantum Electronics*, vol. 50, n° 1, jan. 2018.
- [5] W. Hardle et L. Simar, « Applied Multivariate Statistical Analysis », *Technometrics*, vol. 47, 2005.
- [6] J.-F. Roux, F. Aquistapace, F. Garet, L. Duvillearet and J.-L. Coutaz, « Grating-assisted coupling of terahertz waves into a dielectric waveguide studied by terahertz time-domain spectroscopy », *Applied Optics*, vol. 41, n° 30, p. 6507, oct. 2002.
- [7] F. Bonnefoy *et al.*, « Authentication in the THz domain: a new tool to fight counterfeiting », *9th THz Days Conference, Dunkirk, France*, 2017.
- [8] Z. Ali *et al.*, « Potential of chipless authentication based on randomness inherent in fabrication process for RF and THz », *11th European Conference on Antennas and Propagation (EUCAP)*, Paris, France, p. 2559–2563, 2017.
- [9] C. E. Rasmussen and C. K. I. Williams, « Gaussian Processes for Machine Learning », *the MIT Press*, 2006.
- [10] A.M. Martinez and A.C. Kak, « PCA versus LDA », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, n° 23, p. 228–233, 2001.