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:


HAL Id: hal-02282841
https://hal.archives-ouvertes.fr/hal-02282841
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. Salhi1,2, F. Bonnefoy1, S. Girard3, M. Bernier1, N. Barbot1, R. Siragusa3, E. Perret1 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 $\Gamma = 800 \, \mu m$, groove’s depth $p = 280 \, \mu m$, card thickness $e_{\text{card}} = 775 \, \mu m$, card refractive index $n_{\text{card}} = 1.73$ and absorption coefficient $\alpha = 17 \, \text{cm}^{-1}$ at 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 unitarily 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 two twin tags, each containing 70 frequency points, as a statistical dataset thereby composed of 3.08 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 $\varepsilon^{k,FN}$ to wrongly authenticate the tag $k$ as the tag $t$ (False-Negative case, FN), is directly linked to the clusters size and proximity. $\varepsilon^{k,FN}$ is numerically calculated from the $N$ measurements of the $k^\text{th}$ tag using:

$$
\varepsilon^{k,FN} = \frac{1}{N} \sum_{l=1}^{N} \mathbbm{1}(f_k(x_l) \leq f_t(x_l)), \quad x_l \sim f_k(x)
$$

(1)
With \( f_k(x_i) \): 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:

\[
E_k = \frac{1}{M} \sum_{t=1, t \neq k}^{M} \frac{F(t)}{N}
\]

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).](image)

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 can reach 99.994% using only the 3 first axis from LDA. 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**


