

Open Research Online

The Open University's repository of research publications and other research outputs

Dispersion-Independent Terahertz Classification Based on Geometric Algebra for Substance Detection

Conference or Workshop Item

How to cite:

Zhou, S. L.; Valchev, D. G.; Dinovitser, A.; Chappell, J. M.; Iqbal, A.; Ng, B. W-H.; Kee, T. W. and Abbott, D. (2016). Dispersion-Independent Terahertz Classification Based on Geometric Algebra for Substance Detection. In: 41st International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz), 25-30 Sep 2016, Copenhagen, Denmark.

For guidance on citations see [FAQs](#).

© 2016 IEEE

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1109/IRMMW-THz.2016.7758626>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Dispersion-Independent Terahertz Classification Based on Geometric Algebra for Substance Detection

S. L. Zhou* D. G. Valchev[†] A. Dinovits[‡] J. M. Chappell[‡] A. Iqbal[‡] B. W.-H. Ng[‡] T. W. Kee[§] and D. Abbott[‡]

*School of Engineering and Technology, Southwest University, China

[†]Department of Computer Systems and Technologies, University of Food Technologies, Plovdiv, Bulgaria

[‡] School of Electrical and Electronic Engineering, The University of Adelaide, Australia

[§]School of Chemistry, The University of Adelaide, Australia

Abstract—We demonstrate and validate Geometric Algebra (GA) based terahertz (THz) signal classification of various powders in tablet form of various thicknesses, and compare the results with a conventional Support Vector Machine (SVM) approach. By using geometric algebra we can perform classification independently of dispersion and hence independently of the transmission path length through the sample. In principle, it may be possible to extend the GA coordinate-free transformation to other types of pulsed signals, such as pulsed microwaves or even acoustic signals in such fields as seismology. The classifier is available for download at Github, <https://github.com/swuzhou/slshengling-zhou/blob/geometric-algebra-classifier/GAclassifier/>.

I. INTRODUCTION AND BACKGROUND

WITH terahertz time-domain spectroscopy (THz-TDS) systems, different substances can be distinguished via adoption of a signal classifier that is trained to discriminate absorption spectra in the terahertz regime [1], [2]. However, a problem with using pulsed THz-TDS systems is the dispersion of the terahertz pulse through the sample [3]. This dispersion results in temporal pulse spreading and means that, if a signal classifier has been trained using both frequency and phase data for a sample of one thickness, it may not always work for an arbitrary sample thickness.

To solve this problem we introduce theory based on geometric algebra (GA) and show how it may be used for classifying the terahertz signal, isolating the influence of dispersion. We define a conversion function $g(\cdot)$ which maps a m -dimensional complex transfer function vector \tilde{h} onto a $2m$ -dimensional real vector \mathbf{h} . The function is defined as follows:

$$\mathbf{h} = g(\tilde{h}) = \sum_{i=1}^m \angle \tilde{h} e_i + \sum_{i=1}^m \ln |\tilde{h}| e_{i+m}. \quad (1)$$

It is demonstrated that all vectors \mathbf{h} corresponding to samples of one substance k , so that with different thickness lie on the unique plane \mathbf{H}_k ,

$$\mathbf{H}_k = \frac{\mathbf{h}_k^{d_1} \wedge \mathbf{h}_k^{d_2}}{|\mathbf{h}_k^{d_1} \wedge \mathbf{h}_k^{d_2}|} \quad (2)$$

where $\mathbf{h}_k^{d_1}$ and $\mathbf{h}_k^{d_2}$ denote corresponding transfer function vectors gained from samples of the substance k with thickness d_1 and d_2 separately. Once the set of all planes for the different substances under investigation is determined, the measurement of the THz spectrum of an unknown substance x with unknown sample thickness can take place, producing a measurement vector \mathbf{h}_x . The wedge product $\lambda_k^x = \left| \frac{\mathbf{h}_x}{|\mathbf{h}_x|} \wedge \mathbf{H}_k \right|$ is defined as the criterion for substance identification. The lower the value λ_k^x the more likely the measured unknown substance x with unknown sample thickness is the substance corresponding to that plane.

Therefore, a substance identification method based on minimum value of the parameter λ is presented as:

$$x = \arg \min_k \{\lambda_k^x\} = \arg \min_k \left| \frac{\mathbf{h}_x}{|\mathbf{h}_x|} \wedge \mathbf{H}_k \right|, k = 1, 2, \dots, N. \quad (3)$$

The plane with the minimum value of λ with the measurement vector \mathbf{h}_x is most likely the plane of the unknown measured substance. This algorithm is illustrated in Fig. 1.

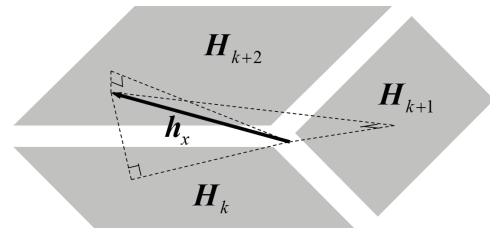


Fig. 1. Illustration of the terahertz signal classification. The closer the measurement vector \mathbf{h}_x is to a particular plane, the lower the value is of its wedge product with that plane and the more likely the measured unknown substance with unknown sample thickness is the substance corresponding to that plane.

II. RESULTS

To systematically evaluate the performance of the proposed GA theory, THz frequency domain spectra from all four substances, melamine, tartaric acid, lactose, and glucose of five different thicknesses, 1.0 mm, 1.5 mm, 2.0 mm, 2.5 mm

TABLE I
CLASSIFICATION PERFORMANCE (%) VS SUBSTANCE THICKNESS USING SVM AND GA CLASSIFIERS.

Substances	SVM					GA				
	1.0 mm	1.5 mm	2.0 mm	2.5 mm	3.0 mm	1.0 mm	1.5 mm	2.0 mm	2.5 mm	3.0 mm
Melamine	54.00	100.00	100.00	98.25	93.25	82.50	97.75	100.00	99.75	99.75
Tartaric Acid	100.00	100.00	100.00	100.00	85.25	100.00	99.75	99.75	99.50	99.50
Lactose	99.00	99.50	100.00	99.00	62.70	99.75	98.25	100.00	100.00	100.00
Glucose	59.25	100.00	100.00	84.00	80.50	100.00	97.50	97.75	86.50	85.00
Overall	78.06	99.87	100.00	95.31	80.43	95.56	98.31	99.37	96.43	96.06

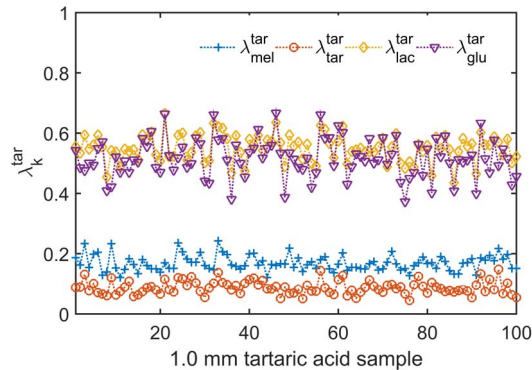


Fig. 2. Illustration of substance identification based on the GA classifier. Here, 100 tartaric acid vectors are obtained from 1.0 mm data samples. Their *vector rejection* magnitudes λ_k^{tar} to each plane are calculated. Here, λ_k^{tar} denotes the magnitude of the tartaric acid vector rejection to plane k , where k corresponds to four substances, melamine, tartaric acid, lactose and glucose. In this graph, all data points from a 1.0 mm tartaric acid sample are classified as tartaric acid correctly by the GA classifier.

and 3.0 mm are used. The classification task is to correctly identify the specific tablet sample given unknown thickness.

First, the complex transfer function vectors are collected, then we calculate their corresponding vectors by via Eqn. 1. For each substance, one vector from a 1.5 mm and a 2.0 mm tablet sample are randomly selected and form the unit plane according to Eqn. 2. For any vector \mathbf{h}_x^d corresponding to a signal from an unknown substance x with unknown thickness d to be identified, we computed the value of $\left| \frac{\mathbf{h}_x^d}{|\mathbf{h}_x^d|} \wedge \mathbf{H}_k \right|$, which is used as the criterion for substance identification. Then the unknown sample is identified as the k -th substance based on the minimum magnitude.

Fig.2 illustrates how the GA classifier works based on a minimal value of λ . As we can see clearly from this graph, due to inevitable influences of noise and other non-idealities in THz-TDS systems, the tartaric acid vectors obtained from 1.0 mm data samples deviate from their real orientation. Their corresponding $\lambda_{\text{tar}}^{\text{tar}}$ is not the zero but the value of less than 0.15. However, compared to the planes of the other substances the tartaric acid plane always has the lowest λ .

In order to further demonstrate the validity and stability

of the proposed GA classifier, the results are compared with the Gaussian kernel SVM classifier, while the SVM classifier is trained by amplitude spectra of 1.5 mm and 2.0 mm data samples.

Table I reports the classification outcomes both of the SVM classifier and the GA classifier for all four substances at thicknesses of 1.0 mm, 1.5 mm, 2.0 mm, 2.5 mm, and 3.0 mm. As can be seen from this table, due to the influence of dispersion and other inevitable noise, the prediction accuracy of both the GA and SVM classifiers are affected. Generally speaking, the dispersion due to thicker samples appears to have less effect on the GA classifier compared to the SVM classifier. For the SVM classifier trained on half of dataset from substances of 1.5 mm and 2.0 mm thickness, the overall classification accuracy is 99.87% for a substance thickness of 1.5 mm and 100.00% for a substance thickness of 2.0 mm. Notice that the overall classification accuracy of the SVM drops to as low as 54.00% for thicknesses that are outside the training set. While the overall classification accuracy of GA classifier remains above 96.06%. Even in the worst case, (melamine at 3.0 mm), the GA prediction accuracy is still above 82.00%.

III. CONCLUSION

Due to the presence of unquantifiable dispersion, classification tasks are typically difficult to perform on samples having non-uniform thickness or unknown thickness. Our work here now solves this problem. The results confirm the superiority in classification accuracy and robustness of our GA-based classifier. It may also be concluded that the GA classifier is a more powerful and less complex algorithm, without the need for any tuning parameters.

REFERENCES

- [1] D. Abbott and X. C. Zhang, "Scanning the issue: T-ray imaging, sensing, and detection," *Proceedings of the IEEE*, vol. 95, no. 8, pp. 1509–1513, 2007.
- [2] X. X. Yin, B. W.-H. Ng, and D. Abbott, *Terahertz Imaging for Biomedical Applications: Pattern Recognition and Tomographic Reconstruction*, Springer, 2012.
- [3] K. Kawase, Y. Ogawa, Y. Watanabe, and H. Inoue, "Non-destructive terahertz imaging of illicit drugs using spectral fingerprints," *Opt. Express*, vol. 11, pp. 2549–2554, 2003.