Pattern Recognition with Machine Learning on Terahertz Images

FIR Center of Univ. of Fukui[1], National Institute for Materials Science[2], ○Dmitry Bulgarevich[1,2], Hideaki Kitahara[1], Masahiro Kusano[2], Takashi Furuya[1], Jessica Afalla[1], Valynn Mag-usara[1], Masahiko Tani[1], and Makoto Watanabe[2]

E-mail: bulgarevich@fir.u-fukui.ac.jp

The terahertz time-domain spectroscopy (THz-TDS) imaging is inherently the hyperspectral technique with sliced imaging in frequency/time domains. It has a promising potential for non-destructive testing (NDT) of various materials. However, the associated large image data volumes and complex image contrasts could make the analysis and interpretation very difficult and time consuming. In recent years, the dramatic progress had been made in automated image pattern recognition techniques for automobile, medical [1], biological, agricultural, IT, security, etc. applications. With suitable algorithm, image database, and powerful computers, the large volumes of image data could be classified in a short time with high accuracy. In this respect, we studied the feasibility of pattern recognition on THz-TDS images of rusted steel with Random Forest machine learning algorithm [2]. It was found that for single hyperspectral set of THz images, the Classifier can be created with out-of-bag (OOB) error below 1 % and reasonable segmentation results (see Fig. 1). For time-apart measured image datasets, the standardized procedure of image preprocessing was necessary to create/apply the single Classifier. Its usage for different sets was limited to frequencies of 1±0.2 THz. More advanced image preprocessing or/and Random Forest code is necessary to improve the Classifier robustness. Here, it should be stressed that application of such technique is novel for THz-TDS imaging, NDT, and materials science fields. In principle, much more complex patterns could be auto-classified/analyzed compared to ones in Fig. 1.

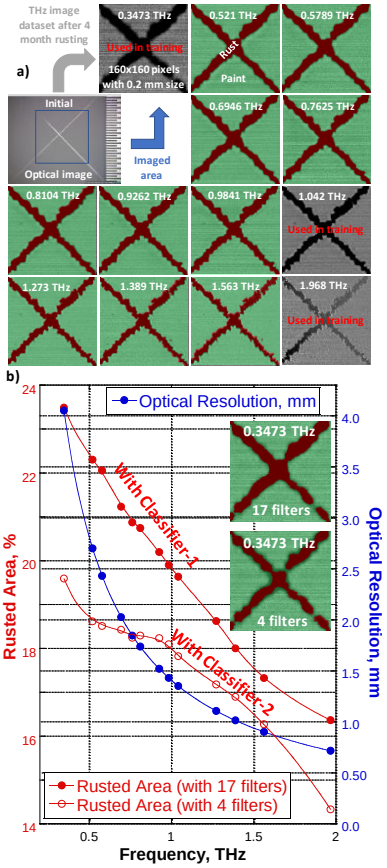

Figure 1. a) The overlays of reflection THz images of painted, cross-scratched, and subsequently rusted steel sample with auto-segmented ones into rusted and painted regions. b) The Classifier quality depends on user-supervised training for lower OOB error and better overlays and correlation with optical resolution.

1. Oliveira, R. B., *et al*. Computational methods for the image segmentation of pigmented skin lesions: a review. *Comput. Methods Programs Biomed*. **131**, 127-141 (2016).

2. Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).

© 2018年 応用物理学会