2023 IMEKO TC-4 International Conference on
Metrology for Archaeology and Cultural Heritage
Rome, Italy, October 19-21, 2023

Terahertz Identification of Characters Written in Iron-Gall Ink on Stacked Paper

Haolian Shi[1,2], Leor Jacobi[3], Alexandre Locquet[2,1], D. S. Citrin[1,2]

[1] *School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 20223-0250 USA*

[2] *Georgia Tech-CNRS IRL2958, Georgia Tech-Europe, 2 Rue Marconi, 57070 Metz, France, david.citrin@ece.gatech.edu*

[3] *Device Spectroscopy Laboratory, Institute for Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan 5290002, Israel*

***Abstract* – Iron-gall ink has been widely used ink in the production of documents and artworks since the fifth century C.E. Terahertz imaging is used to read characters written in such ink on stacked sheets of paper. Several contrast mechanisms including maximum/minimum of the reflected terahertz signal, signature frequency, Gaussian mixture model, and short-time Fourier transform are tested to compare the image quality of the buried text. The results show that the short-time Fourier transform mechanism outperforms other contrast mechanism and has the ability to extract clear images of buried text within paper stacks.**

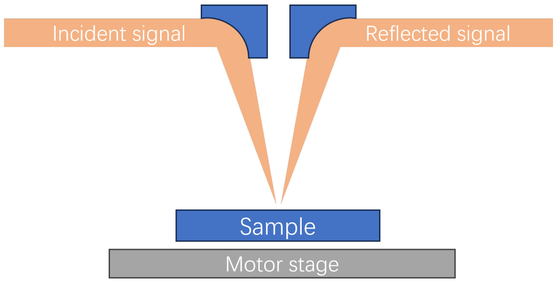

Fig. 1. Setup of the measurement

I. INTRODUCTION

Iron-gall (IG) ink has been widely used since the fifth century C.E. [1]. The chemistry of IG ink is discussed in Ref. [2]. Terahertz imaging of various inks has been investigated in a number of studies [3, 4]. However, imaging text written paper stacks of more than one sheet is not common [5, 6].

Terahertz imaging is a nondestructive, noninvasive, and nonionizing imaging technique that can be applied to a wide variety of materials, including the ink found in cultural-heritage artifacts. Of specific interest is to extract text from stacks of paper sheets without separating them, as it may not be possible to separate sheets of paper without casing them irreversible damage as shown in [7]. Terahertz imaging may provide the capability to read text hidden in stacked paper, but to obtain high-quality images, it is of importance to establish contrast mechanisms that yield the desired information. In this work, similar to [7], paper stacked are studied however with a specific ink of IG. We extend the discussion of contrast mechanisms to explore a suitable one for imaging. Several contrast mechanisms based on reflective terahertz imaging, both in time and frequency domains, are compared focusing in IG ink on paper (standard copy paper). The terahertz reflectivities of IG ink and paper are different, which suggests using the maximum or minimum of the reflected signal a potential way to distinguish ink from paper. Moreover, in the frequency domain, there may exist spectroscopic signatures for ink or paper that can be used as the basis of a contrast mechanism. Beyond simple time- or frequency-domain-based contrast mechanisms, more sophisticated statistical analyses of reflected terahertz signals are also of interest. We therefore employ a Gaussian mixture model (GMM). Finally, a short-time Fourier transform (STFT)-based contrast mechanism is introduced, the method combining the advantages associated with time- and frequency-domain-based contrast mechanisms. This method can obtain the location of text written in IG ink in a paper stack by using time windowing as well as the strong contrast of the image by utilizing frequency information in each time window.

A TPS Spectra 3000 (TeraView Ltd., Cambridge UK) is used to measure the reflected signal. The set up of experiment is shown in Figure 1. The generated terahertz pulse is directed towards the sample using mirrors and other optics. The intensity and phase of the reflected pulse provide information about the sample's properties.

II. RESULT AND DISCUSSION

A. *One sheet of paper with characters written on both sides*

Figure 2 is a schematic of the sample studied first: a single paper sheet with characters "2" and "H" written in iron-gall ink on opposite sides of the paper. "H" is on the side of the paper from which the terahertz signal is incident.

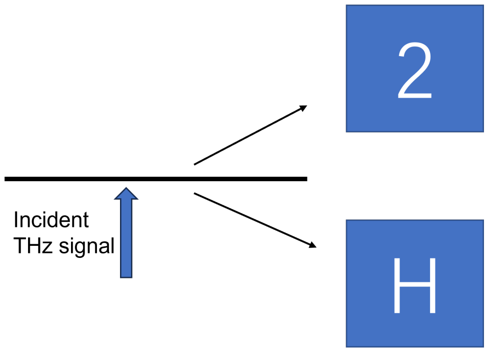

Fig. 2. Schematic of paper sheet with characters "2" and "H" written on opposite sides of the paper. The incident direction is from 'H' side to '2' side

We can classify four different regions: (1) that without any portion of a character present, (2) that with a portion of the character "H," (3) that with a portion of the character "2,", and (4) that with portions of both "H" and "2."

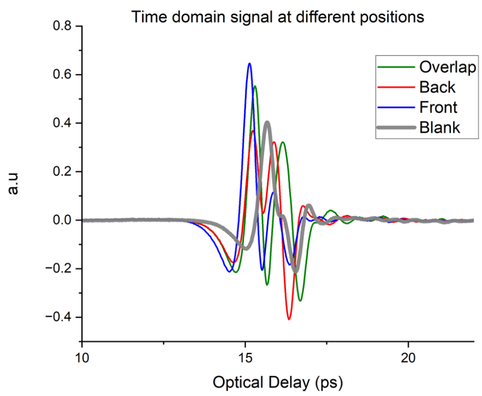

Fig. 3. Reflected time-domain THz signals at different positions on the sheet of paper corresponding to regions (1)-(4).

Examples of the reflected terahertz signals characteristic of the four regions, are shown in Fig. 3. The gray curve shows the reflected time-domain terahertz signal in region (i). It can be seen clearly that there is only on sheet of paper in the signal as evidenced by two reflected terahertz pulses: one from the air/paper interface (on the side of incidence) and one by the paper/air interface (on the opposite side). The former/latter has the same/opposite polarity of the incident pulse due to the opposite signs of the Fresnel reflection coefficients for the two interfaces. The blue curve shows a typical reflection signal in region (2). It can be seen clearly that there is a strong reflection caused by the ink on the side of incidence. An example of the reflected terahertz signal in region (3), where the character is

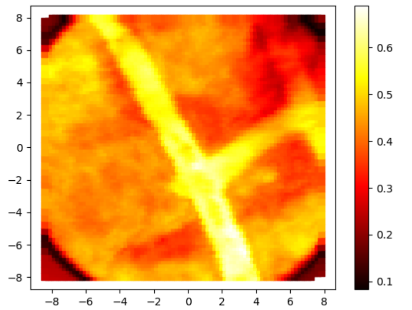

Fig. 4. C-scan of the sheet of paper in the vicinity of the characters. The color in each pixel is the maximum amplitude of the reflected time-domain terahertz signal at a given position.

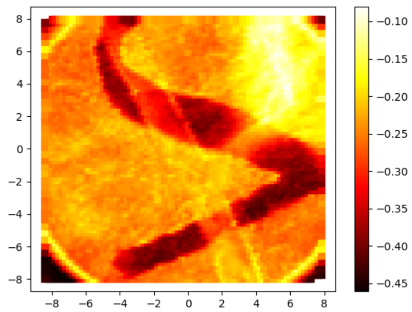

Fig. 5. C-scan of the sheet of paper in the vicinity of the characters. The color in each pixel is the minimum amplitude of the reflected time-domain terahertz signal at a given position.

present on the opposite side, is given by the red curve. The first reflection from the air/paper interface is weaker than in region (2), but the strength of reflection from the back paper/air interface is relatively strongly enhanced. For region (4), the green curve, both the first and second reflections are strong. Based on these results, the strengths of the various reflections are a good candidate for determining which region corresponds to a measurement made on a given point on the sheet of paper.

Figure 4 shows a C-scan of the sheet of paper in the vicinity of the characters; the color in each pixel is the maximum amplitude of the reflected time-domain terahertz signal at a given position, while in Fig. 5, a C-scan based on the signal minimum is plotted. These C-scans can easily be understood on the basis of our discussion of Fig. 3. We see both the "H" and the "2," and can distinguish from these two C-scans taken together which side of the sheet of paper each character is on.

We next consider a contrast mechanism based on a frequency-domain analysis of the reflected terahertz signals. Figure 6 shows the amplitude of the Fourier transform of reflected signals in regions (1)-(4). For region (1), there is only one pronounced dip (absorption line) at a frequency ∼1.1 THz. For regions (2)-(4), however, this dip splits into a doublet, leaving no dip at ∼1.1 THz. Further features are seen distinguishing regions (2), (3), and (4). As is show in Fig 7, it is clear that the image can clearly distinguish paper and ink with good contrast by using the signal amplitude at 1.1 THz. To determine on which side a character is written (or both) requires a more sophisticated approach.

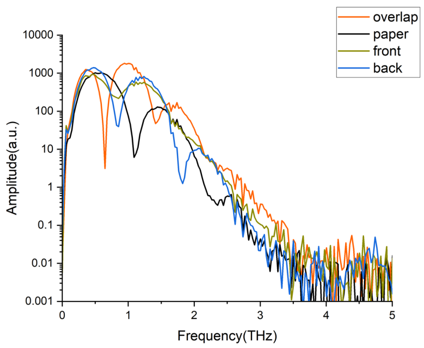

Fig. 6. Fourier transform of reflected THz signals at different positions on the sheet of paper corresponding to regions (1)-(4). See text for definitions of these regions.

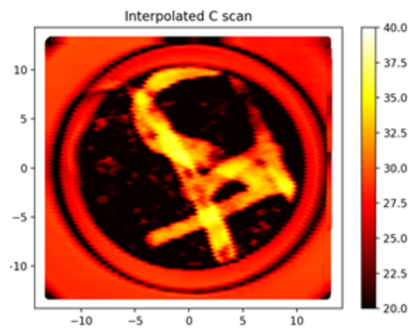

Fig. 7. C-scan of the sheet of paper in the vicinity of the characters. The color in each pixel is the amplitude of the Fourier transform of the reflected THz signal at a given position at frequency 1.1 THz.

Because several features are needed to distinguish the four regions, we next employ a Gaussian mixture model (GMM) to group the signals into distinct clusters to pigeonhole each pixel into the four classifications [8]. We set the cluster numner at 4, which is sensible as we seek to indentify four distinct regions. The result can be seen in Fig 8. The four clusters are mapped to different colors enabling one to easily distinguish on which side of the paper a character, if present, is written. The approach, moreover, enables one to clearly identify region (4) in which the two characters on opposite sides of the paper overlap.

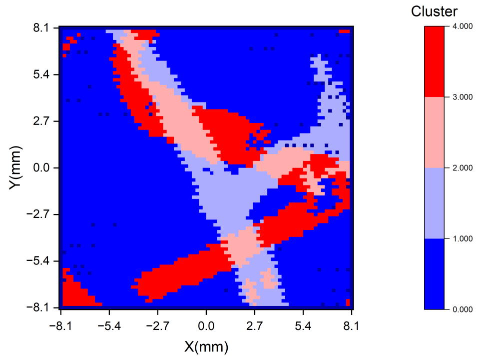

Fig. 8. C-scan based with contrast mechanism based on GMM with cluster number set to 4.

B. Two sheets of paper with characters written on one side of each sheet

Next, we consider two sheets of paper. The sheet on the side of incidence (the reverse side) has the character "H" ("2") written on the obverse of the sheet in iron-gall ink; Fig. 9 is a schematic diagram. There are total two layers of paper and with ink on the side that is facing the signal on each paper.

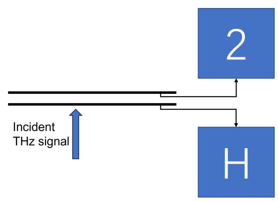

Fig. 9. Schematic of two paper sheets with characters "H" and "2" written on obverse sides of the paper sheets. The incident direction is from 'H' side to '2' side

C-scans with contrast mechanism based on the maximum and minimum of the reflected time-domain terahertz signals are shown in Figs. 10 and 11, respectively. Although the character "H" on the first sheet of paper encountered by the incident terahertz signal is easy to distinguish in the C-scan based on the signal maximum, the "2" on the second layer is indistinct. Because both characters are on the obverse side, they both correspond to relative maxima in the signal; however, due to attenuation in propagating through two sheets of paper the reflection from the second sheet of paper is much less, i.e. the signature of the "2" is weak. As a result, only use the maximum of the signals as a contrast mechanism is not enough in this case. Also note

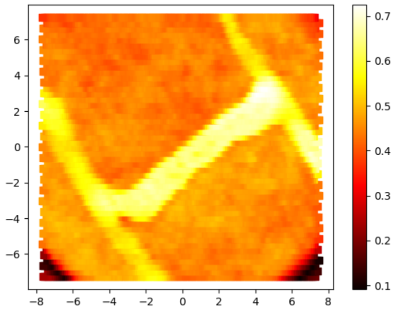

Fig. 10. C-scan of two sheets of paper in the vicinity of the characters. The color in each pixel is the maximum amplitude of the reflected time-domain terahertz signal at a given position.

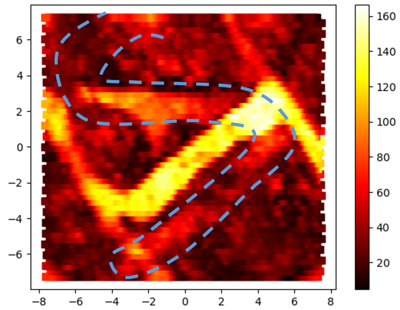

Fig. 12. C-scan of the two sheets of paper in the vicinity of the characters. The color in each pixel is the amplitude of the Fourier transform of the reflected THz signal at a given position at frequency 1.1 THz. The character "2" is outlined as a guide to the eye.

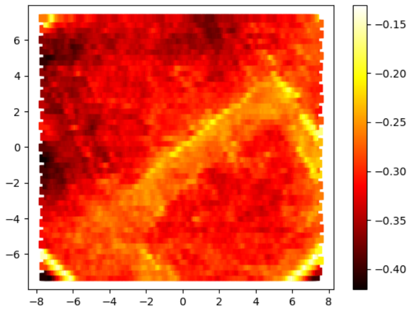

Fig. 11. C-scan of two sheets of paper in the vicinity of the characters. The color in each pixel is the minimum amplitude of the reflected time-domain terahertz signal at a given position.

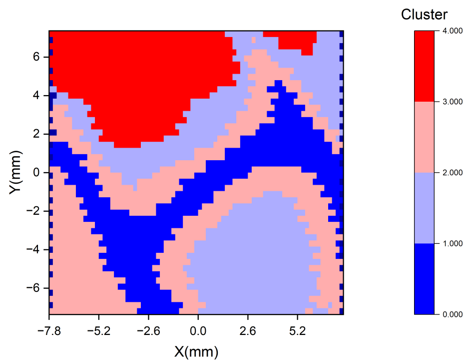

Fig. 13. C-scan based with contrast mechanism based on GMM with cluster number set to 4.

that the contrast mechanism based on the minimum in this case is of little help. Serrated edges can be observed in the figure is mainly caused by the bidirectional rotation errors of the motor. During the scan, the motor would rotate in two directions and lead to the unalignment of the measure point. This effect can be suppressed by applying smaller step size.

In the frequency domain, Fig. 12 shows the C-scan at 1.1 THz. We can distinguish both the "H" and the "2" (the latter outlined as a guide to the eye), though some work is still needed to enhance the contrast.

Figure 13 is a C-scan based on a GMM analysis with cluster number 4 as in the one sheet case. While the character "H" on the first sheet of paper is clearly seen, the "2" on the second sheet is not.

C. *Short-time Fourier transform based imaging*

Clearly, the location of the characters is contained in the respective time slice of the signal while the presence of iron-gall ink was also seen in the frequency domain. This suggests that a properly chosen spectrogram of the reflected terahertz signal might provide better contrast for the two sheets of paper. We therefore propose a short time Fourier transform (STFT). In each time window chosen, the Fourier transform of the reflected signal is performed. Figure 14 is a an STFT spectrogram for an example reflected time-domain terahertz signal. The window is 1 ps which is roughly the duration of the incident terahertz pulse. It can be seen that at positions where the reflections occur, there will be strong frequency response.

By selecting the time windows in the time domain around the reflection from the first sheet and integrating over frequency, we obtain an image localized in the first sheet of paper, as shown in Fig. 15. Similarly, we can select the time window corresponding to the second sheet of paper, as shown in Fig. 16. We note that this approach succeeds in distinguishing the characters written on the individual sheets of paper.

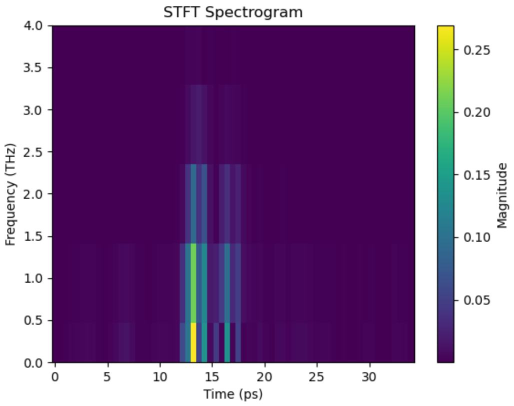

Fig. 14. Example of STFT (see main text) of reflected signals.

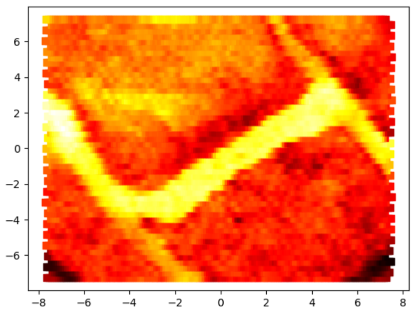

Fig. 15. C-scan with STFT-based contrast mechanism to provide information about the first sheet of paper. The character "H" is seen.

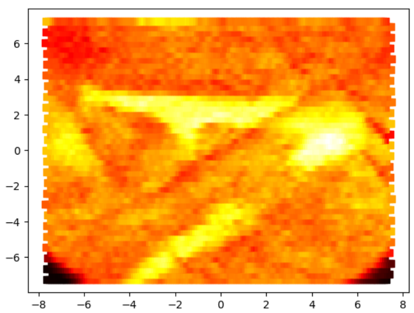

Fig. 16. C-scan with STFT-based contrast mechanism to provide information about the second sheet of paper. The character "2" is seen.

III. CONCLUSION

This study present a comparison between a variety of contrast mechanisms employed for imaging characters written in iron-gall ink on paper, both for one and two paper sheets. We found out that Hadar ink has a strong reflection to THz in the band below 2 THz which facilitate the imaging of the ink. While C-scans based on the maximum or minimum of the reflected terahertz signal were shown to be suitable for one paper sheet, the approach did not perform satisfactorily on stacks of two sheets of paper. A simple frequency-domain-based contrast mechanism has the ability to distinguish area on the paper with and without ink; however, it lacks the ability to distinguish between layers. A Gaussian mixture model approach was able to cluster signals for one sheet of paper; however, it has limitations for the paper stacks. An short-time Fourier transform-based contrast mechanism was then explored to image paper stacks. This approach identified characters written on two sheets of paper and distinguished on which sheet each was written. Clearly, these approaches merit study to be extended to treat stacks with multiple sheets of paper with characters written on both sides.

IV. ACKNOWLEDGEMENT

LJ acknowledges the financial support of the Alexander von Humboldt Stiftung.

REFERENCES

[1] M. T. Ruggiero, T. Bardon, M. Strlič, P. F. Taday, and T. M. Korter, "The role of terahertz polariton absorption in the characterization of crystalline iron sulfate hydrates", Phys. Chem. Chem. Phys. , vol. 17, No. 14, 2015, pp. 9326-9334.

[2] E. Eusman, "Iron gall ink - Chemistry", The Iron Gall Ink Website,1998.https://web.archive.org/web/20190424073116/https://irongallink.org/igi_indexedde.html

[3] T. Bardon, R. K. May, J. Jackson, et al., "Contrast in Terahertz Images of Archival Documents-Part I: Influence of the Optical Parameters from the Ink and Support", J Infrared Milli Terahz Waves, vol. 38, 2017, pp. 443-466.

[4] J. Tasseva, A. Taschin, P. Bartolini, J. Striova, R. Fontana, and R. Torre, "Thin layered drawing media probed by THz time-domain spectroscopy", Analyst, vol. 142, No. 1, 2017, pp. 42-47.

[5] J.E. Ensley, K.H. Tachau, S.A. Walsh, "Fragment of lost 12th-century epic poem found in another book's binding", The Guardian, 18 November 2021.

[6] H. Zhang, G. Simon, L. Moser, J. Atha, P. Dilley, E.A. Hoffman, and M. Sonka, "Using computed tomography to recover hidden medieval fragments beneath early modern leather bindings, first results", Herit Sci.

vol. 11(1), 2023, p. 82. DOI: 10.1186/s40494-023-00912-9.

[7] Redo-Sanchez, A., Heshmat, B., Aghasi, A. et al. “Terahertz time-gated spectral imaging for content extraction through layered structures”, Nat Commun 7, 12665 (2016). https://doi.org/10.1038/ncomms12665

[8] D. Reynolds, “Gaussian Mixture Models," in Encyclopedia of Biometrics, S. Z. Li and A. Jain, Eds., Springer US, Boston, MA, 2009, pp. 659-663. https://doi.org/10.1007/978-0-387-73003-5_196.