

Speech Encryption with Fractional Watermark

Yan Sun^{1,2,*}, Cun Zhu¹ and Qi Cui³

¹Key Laboratory of Artificial Intelligence Application Technology State Ethnic Affairs Commission, Qinghai Minzu University, Xining, 810007, Qinghai, China

²School of Computer of Qinghai Normal University, Xining, 810003, Qinghai, China

³The Department of Electrical and Computer Engineering, University of Windsor, Windsor, N9B 3P4, Canada

*Corresponding Author: Yan Sun. Email: sy0623@163.com

Received: 02 March 2022; Accepted: 12 April 2022

Abstract: Research on the feature of speech and image signals are carried out from two perspectives, the time domain and the frequency domain. The speech and image signals are a non-stationary signal, so FT is not used for the non-stationary characteristics of the signal. When short-term stable speech is obtained by windowing and framing the subsequent processing of the signal is completed by the Discrete Fourier Transform (DFT). The Fast Discrete Fourier Transform is a commonly used analysis method for speech and image signal processing in frequency domain. It has the problem of adjusting window size to a for desired resolution. But the Fractional Fourier Transform can have both time domain and frequency domain processing capabilities. This paper performs global processing speech encryption by combining speech with image of Fractional Fourier Transform. The speech signal is embedded watermark image that is processed by fractional transformation, and the embedded watermark has the effect of rotation and superposition, which improves the security of the speech. The paper results show that the proposed speech encryption method has a higher security level by Fractional Fourier Transform. The technology is easy to extend to practical applications.

Keywords: Fractional Fourier Transform; watermark; speech signal processing; image processing

1 Introduction

Research on the feature of speech signals is carried out from two perspectives, the time domain and the frequency domain. The underlying mathematical tool is the Fourier Transform (FT) in the frequency domain. The speech signal is a non-stationary signal, so FT isn't used for the non-stationary characteristics of the speech. When short-term stable speech is obtained by windowing and framing, the subsequent processing of the speech is completed by the Discrete Fourier Transform (DFT). However, this processing method will have a time-frequency resolution trade-off. It is not possible to see both the whole contour of the speech and the fine structure of the speech signal at the same



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

time domain. For this reason, this work uses the fractional Fourier transform (FRFT) to process the speech signal.

Fractional Fourier transform (FRFT) belongs to the category of fractional calculus. In dealing with fluid physics, there are many achievements in fractional calculus [1]. Intuitively, the physics of speech can be analogous to fluid. Therefore, this article uses the FRFT to conduct a preliminary study on speech processing with Fractional watermark.

2 Fractional Fourier Transform

2.1 Fractional Fourier Development Background

In 1929, Wiener started the earliest research work on the fractional Fourier transform [2]. Mermelstein et al. [3] formally proposed the fractional Fourier transform method in 1980, starting from the polynomial of the Fourier series, calculating its eigenvalues and eigen functions [4]. In 1993, McBride et al. used the integral form a more rigorous mathematical definition. The optical verification of the Fractional Fourier transform is realized by Mendlovic et al. [4]. A reasonable physical explanation is given by the experimental results. The fractional Fourier transform has been widely used in the field of optics [5].

Unfortunately, due to the lack of effective physical explanations and fast algorithms in the field of signal processing. The progress had been slow. In the same year, Almeida pointed out that the fractional Fourier transform can be understood as a rotation in the time-frequency domain. Based on Almeida's ideas, Ozaktas et al. proposed a discrete algorithm with a calculation cost amount equivalent to FFT in 1996. Since then the fractional Fourier transform received the attention in the field of signal processing [6, 7].

The basis of the fractional Fourier transform is fractional calculus. Fractional calculus has been apply to the study of fluid phenomena. In the field of fluid research in China, the processing method of fractional calculus is used. Professor Chen Wen from Hohai University achieved remarked research results [8]. At present, the application of the fractional Fourier transform in the field of speech signals is mainly in the detection of chirp signals, signal reconstruction, and acoustic signal analysis.

2.2 Definition of Fractional Fourier Transform

Fractional Fourier Transform can be defined as follows:

$$\begin{aligned} X_a(u) &= F^p[x(t)] = \int_{-\infty}^{\infty} x(t)K_\alpha(t, u)dt \\ X_p(u) &= F_p[x(t)] = \int_{-\infty}^{\infty} K_p(u, t)x(t)dt \end{aligned} \quad (1)$$

where $F[\cdot]$ represents the FRFT operator, p is a real number, called the order, and let $\alpha = \frac{p\pi}{2}$, as the transformation angle, which is the rotation angle of the speech signal in the time-frequency domain.

$$K_\alpha(t, u) = \begin{cases} \sqrt{\frac{1-j\cot\alpha}{2\pi}} \cdot \exp\left(j\frac{t^2+u^2}{2}\cot\alpha - tu\csc\alpha\right) & \alpha \neq n\pi \\ \delta(t-u) & \alpha \neq 2n\pi \\ \delta(t+u) & \alpha \neq 2n(2n+1)\pi \end{cases} \quad (2)$$

Analyzing formulas (1) and (2), the fractional Fourier transform has periodicity. When $p = 4n + 1$ is substituted into the transformation angle, the period of the fractional Fourier transform is 4. If the period is 4 for transformation, the fractional Fourier transform is equivalent to the traditional Fourier transform (FT). When $p = 4n - 1$, substituting into the formula (2), the fractional Fourier transform

is equivalent to the traditional inverse Fourier transform (IFT). When $p = 4n$, the fractional Fourier transform is the signal $x(t)$ itself. When $p = 4n + 2$, the Fractional Fourier transform of the signal $x(-t)$.

The FRFT is defined by (2) formula, so that all orders have values. Because of periodicity, when $p = 0$, it is the fractional Fourier transform. When $p = 1$, it is the ordinary Fourier transform (FT). When p is between 0 and 1, it is the transition from the FRFT to ordinary The Fourier transform.

The following explains the meaning of time-frequency domain rotation. In the time-frequency coordinate system, let the x-axis be the time axis and the y-axis be the frequency axis. Assuming that the speech signal at time t is transformed by Fourier transform (FT). If the voice signal is rotated twice in a row, it is equivalent to the inverse Fourier transform. The voice signal is rotated four times in a row without the signal being transformed. If the rotation angle of the speech signal is not an integral multiple of $\pi/2$, then the transformation angle of FRFT corresponds to the rotation angle in the counterclockwise direction of the time-frequency coordinate in the time-frequency domain. Combined with the rotation matrix, the time-domain signal and the frequency-domain signal is connected, which denoted by formula (3):

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{pmatrix} \begin{pmatrix} \tau \\ \zeta \end{pmatrix} \tag{3}$$

where $\begin{pmatrix} \tau \\ \zeta \end{pmatrix}$ is intermediate variable, the $\begin{pmatrix} u \\ v \end{pmatrix}$ is the result of the function after the FRFT transformation with Wigner distribution. After the signal passes via FRFT, it contains both time domain and frequency domain information. It is a joint representation of the time domain and frequency domain. When the rotation angle changes from 0 to 90°, it indicates FRFT gradually changes from time domain analysis to frequency domain analysis

2.3 Properties of Fractional Fourier Transform

The operator F^α is to rotate the function by the angle α , and the operator F^β is to rotate by the angle β in the function. If these two operators act on the same function, the original function is equivalent to the angle rotated $\alpha + \beta$, and the function is equivalent to

- 1) **Linear** $F^{\alpha+\beta} = F^\alpha F^\beta$.
- 2) **Reversibility** $F^p [\sum c_n f_n(u)] = \sum c_n [F^p f_n(u)]$.
- 3) **Unitary** $F^0 = F^{-p+p}, (F^p)^{-1} = F^{-p}$.

Assuming that the function is p times of FRFT, then IFRFT, which is equivalent to p order FRFT of the original function, and then conjugate transpose, that is, it satisfies

$$(F^p)^{-1} = (F^p)^H.$$

4) Commutative law $F^\alpha F^\beta = F^\beta F^\alpha$.

The earliest application of the fractional Fourier transform in physical optics. It was verified by Wright. However, the fractional Fourier transform requires discretization by a computer to process. Now, there are three discrete algorithms, such as, based on sampling, based on feature decomposition, and based on linear combination.

The method was completed by Kranuskas and Pei. They performed fractional sampling in the time domain and frequency domain respectively. Ozaktas establishes a connection between the time domain

and frequency domain fractional sampling points [9]. The eigenvalue-decomposition algorithm is the continuous fractional FRFT. The eigenvalue-decomposition algorithm is expanded with the kernel function spectrum and reconstruct the discrete fractional Fourier transform FRFT matrix. Because it includes the calculation of orthogonal normalized eigenvectors, the amount of calculation cost is huge. Discrete fractional Fourier transform of linear combination expresses any fractional FRFT with fractional orders of 0, 1, 2, 3. The FRFT discretization algorithm of linear combination was developed by the Taylor series formula. It has the least smallest amount of calculation cost. However, the disadvantage is that the error is large.

3 Comparison of Fractional Fourier Transform and Fourier Transform of Speech

In 1807, the French scholar J. Fourier first proposed the Fourier transform. After more than one hundred years of development, it has become the main method of signal processing. Its limitation is that it assumed the signal is stable in short time. To overcome the shortcomings of the Fourier transform, time-frequency-domain analysis methods were proposed. Time-frequency analysis can be divided into two categories. The first category is linear time-frequency: such as Gabor, short-time Fourier and wavelet transform. The second category is bilinear time-frequency analysis, such as Cai et al. [10].

The first type of window function and its width have a great influence on the result. The processing of the time-domain dynamic weakens the change process of the speech signal. The second type of bilinear-time-frequency analysis gets rid of the constraint of the window function. But it introduces the cross-term. The cross-term has great interference on the detection and recognition of the signal. Aiming at the shortcomings of the first type of time-frequency analysis, fractional Fourier transform method was proposed in the last century. It not only eliminates the influence of cross terms, but also has the ability to analyze and process signals from a local and global perspective.

At present, fractional Fourier transform and fractional derivative transform are effective tools in the modeling atmospheric or non-Newtonian fluid turbulence and eddy currents. More specifically, the research of fractional Fourier transform mainly focuses on the turbulence wave source area analysis and chip signal and image research [9].

For speech, Fractional Fourier transform has transferred for processing. The reason is as follows. Human voice is the sound wave generated by the vibration of the object. The fluctuation has perceived after transmitted through the air, liquid, or solid body. Turbulence appears when the voice encounters obstacles during transmission. To take an example, human heart disease will change the blood flow of the heart from laminar flow to turbulent flow. When the doctor hears the heart sound with a stethoscope, it has mixed with the sound produced by the turbulence. The doctor listens and analyzes the patient's heart beat. Long-term diagnosis and treatment experience diagnoses the patient's heart disease. However, this process allows the computer to give an analytical diagnosis, that is, to study and simulate the mechanism of the complex [11] sound produced by heart disease through computers and electronic equipment [12–15], then make corresponding computer diagnosis. A more difficult task is the current domestic medical testing needs. The mathematical basis of computer simulation modeling [16] of turbulent sound signals is the fractional Fourier transform. There are few researches on Fractional Fourier transform of sound in China, so it is very important to study the fractional Fourier transform of speech.

4 Experiments

The experimental signal in this paper is a girl’s voice. The discrete Fourier transform and fractional Fourier transform processing experiment has performed under the matlab2016b platform. The sound and image are plotted during the experimental. The frame length of the experiment is 256, and the noise added is white noise. The experimental results are as follows (Figs. 1–6)

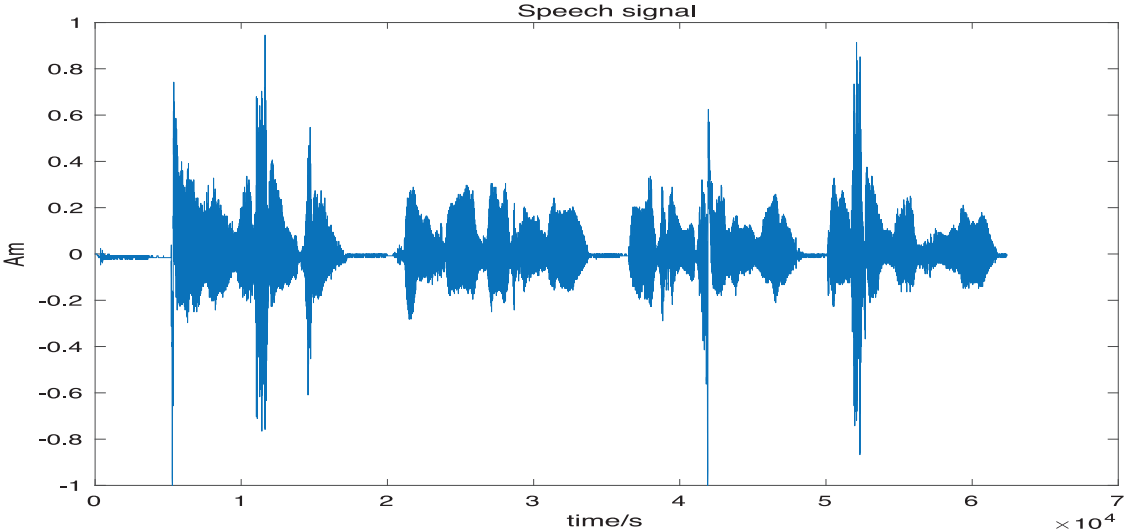


Figure 1: Speech signal

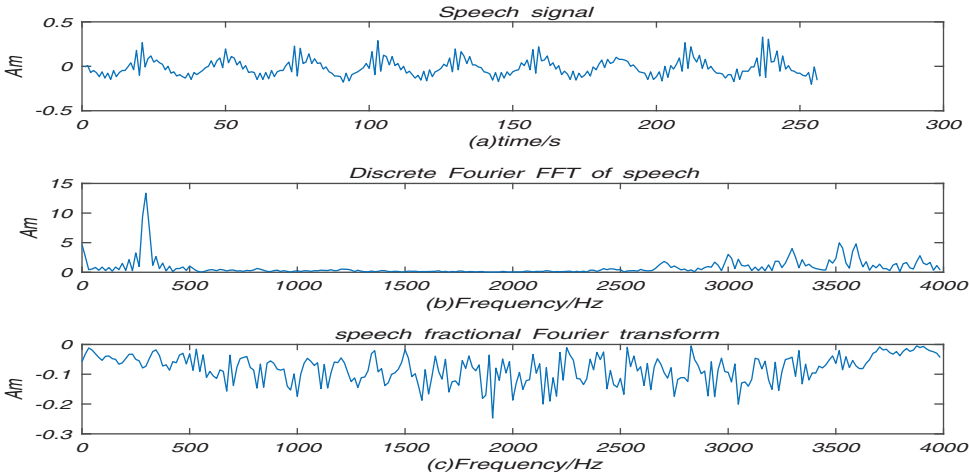


Figure 2: Truncate of speech

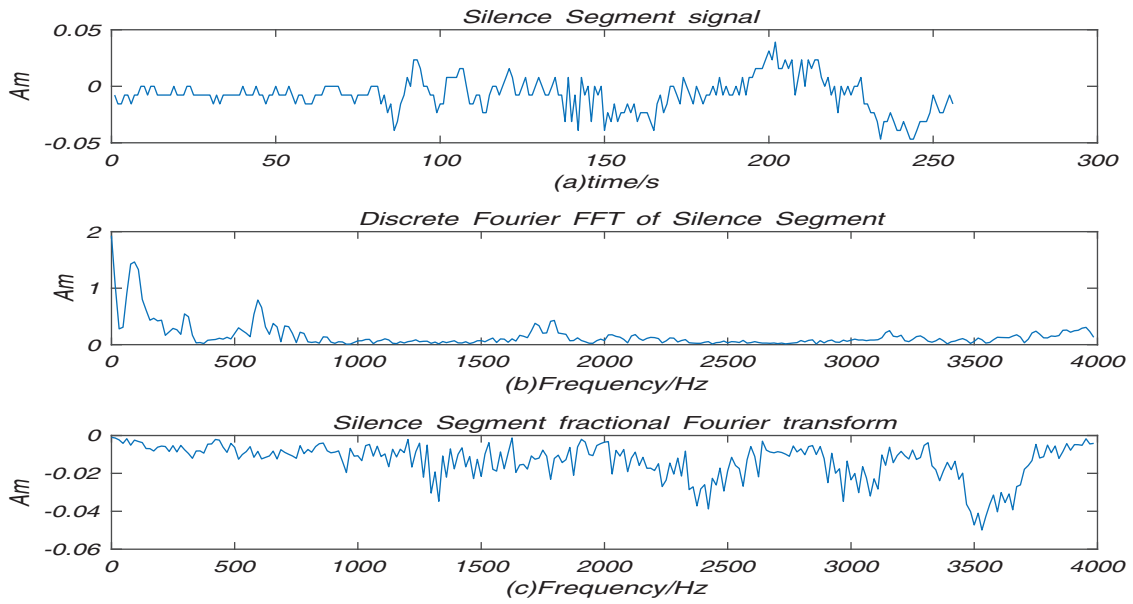


Figure 3: Silence segment of signal

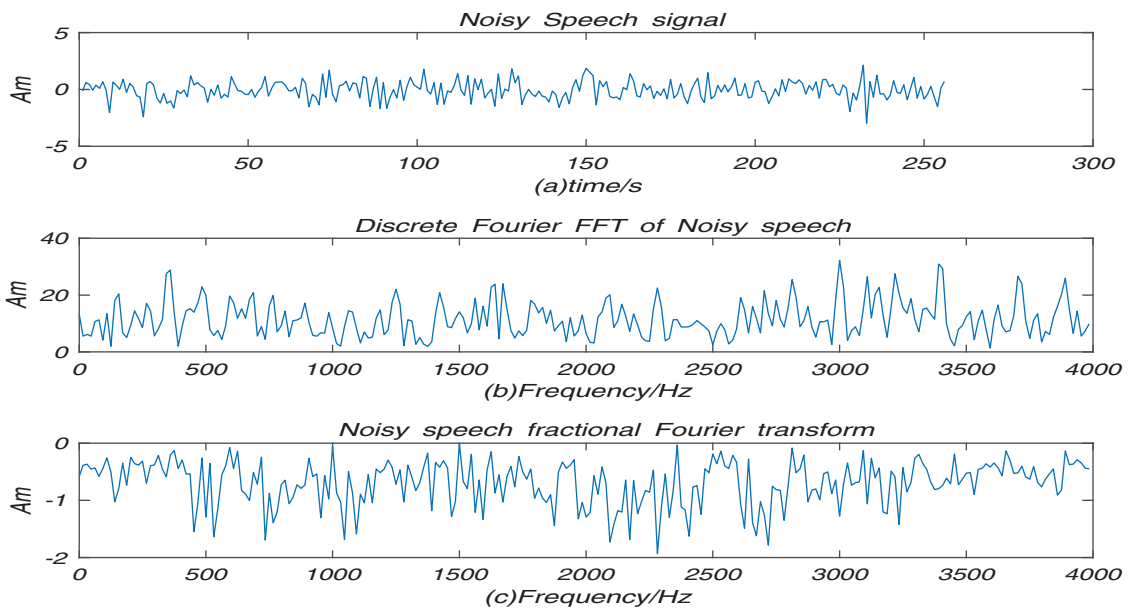


Figure 4: Noisy speech Fourier and fractional Fourier processing



Figure 5: (a) The original image. (b) embedded image

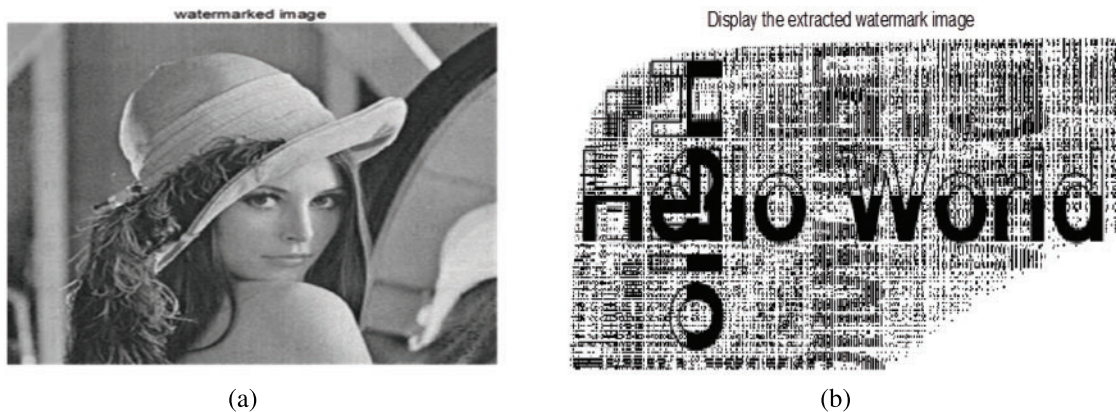


Figure 6: (a) The watermarked image. (b) extracted watermark image

Fig. 1 shows the voice signal waveform with a recording time of more than 6 s, and the content is a five-character Tang poem.

The segment of speech has been processed in the voiced frame in the experimental. The plot (a) of Fig. 2 is the speech signal (b) is the discrete Fourier FFT for speech signal. (c) is the fractional spectrum for speech signal. The waveforms of Figs. 2b and 2c have frequency domain characteristics. It is observed from plot (c) of Fig. 2 that the fractional spectrum has a bilateral spectrum, which is symmetrical around the center point in the Fig. 2. It is symmetrical about the transformation order and the score with the sampling 0 point. In fact, the fractional Fourier transform is unitary. In other words, the result of the speech via fractional Fourier transform, the result in Fig. 2 is symmetric. The speech are symmetrical about the center .

In Fig. 3a is silence signal (b) is the discrete Fourier FFT for silence signal, (c) is the fractional spectrum for silence signal. The waveforms of Figs. 3c and 3d have frequency domain characteristics. The result of Fig. 3 shows speech can be added, and fractional Fourier transform is linear.

Fig. 4 shows noisy speech of the discrete Fourier transform and fractional Fourier transform processing results. Fig. 4a is noisy speech (b) is the discrete Fourier FFT for noisy speech signal. (c) is

the fractional spectrum for noisy speech signal. The result of Fractional Fourier transform can process the speech signal of time-domain and frequency domain.

From the plots in Figs. 2–4, after the fractional Fourier transform of the voiced and unvoiced segments of speech, we understand that these speech segments have the combined characteristics of the time domain and the frequency domain. The speech of fractional Fourier transformed has excellent properties. The properties is linear unitary, reversibility and commutative law. However, from speech of fractional Fourier transform of speech, we cannot be not observed rotational properties of speech.

Below we illustrate this property by performing a fractional transformation on the image.

Fig. 5 shows the original and embedded image used in the experiment. They are processing by gray. (a) is the original lena image. (a) is image of the word Hello World.

In Fig. 6, the result shows the watermarked image and extracted watermark image by fractional watermark [17–20]. From Fig. 6, we can see that the embedded watermark image is processed by fractional transformation, and the embedded watermark has the effect of rotation and superposition. The watermark is added to voice using FFT. It can be embedded into the voice, so that the voice watermark has a more secure performance. In fact, this watermark technology is applied to real world, the image or voice can be better protected in specific use.

5 Conclusion

Our experiments studied the performance of Fractional Fourier Transform (FRFT). We compared with fractional Fourier transform of speech and image. The fractional image is embedded into the fractional speech, which improves the security of the speech. The technology is easy to extend to practical applications.

Acknowledgement: We thank the editors and the anonymous reviewers for their professional and valuable suggestion.

Funding Statement: The work is supported by Regional Innovation Cooperation Project of Sichuan Province (Grant No. 22QYCX0082). Jian-Guo Wei received the grant, and the Science and Technology Plan of Qinghai Province, China (Grant No. 2019-ZJ-7012). Xiu Juan Ma received the grant.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] I. Podlubny, “Geometric and physical interpretation of fractional integral and fractional differentiation,” *Journal of Fractional Calculus & Applied Analysis*, vol. 5, no. 4, pp. 367–386, 2002.
- [2] N. Wiener, “Hermitian polynomials and Fourier analysis,” *J. Math. Phys.*, vol. 8, no. 8, pp. 70–73, 1929.
- [3] S. B. Davis, P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Trans, on Acoustics, Speech & Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.
- [4] D. Mendlovic and H. M. Ozaktas, “Fractional Fourier transforms and their optical implementation: I,” *Journal of the Optical Society of America A*, vol. 10, no. 10, pp. 1875–1881, 1993.
- [5] X. R. Zhang, X. Sun, X. M. Sun, W. Sun and S. K. Jha, “Robust reversible audio watermarking scheme for telemedicine and privacy protection,” *Computers, Materials & Continua*, vol. 71, no. 2, pp. 3035–3050, 2022.

- [6] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. on Speech Audio Process*, vol. 3, pp. 72–83, 1995.
- [7] Martin F. Schwartz and Helen E. Rine, "Identification of speaker sex from isolated, whispered vowels," *Journal of the Acoustical Society of America*, vol. 44, no. 6, pp. 17–36, 1968.
- [8] W. Chen and Y. J. Liang, "Structural derivative based on inverse Mittag-Leffler function for modeling ultraslow diffusion," *Fractional Calculus & Applied Analysis*, vol. 19, no. 5, pp. 1250–1261, 2016.
- [9] H. M. Ozaktas and D. Mendlovic, "Fractional fourier transforms and their optical implementation. II," *Journal of the Optical Society of America A*, no. 10, pp. 2522–2531, 1993.
- [10] P. S. Wright, "Short-time Fourier transforms and Wigner-Ville distributions applied to the calibration of power frequency harmonic analyzers," *IEEE Transactions on Instrumentation & Measurement*, vol. 48, no. 2, pp. 475–478, 1999.
- [11] Y. Sun and H. X. Zhao, "Eigenvalue-based entropy and spectrum of bipartite digraph," *Complex & Intelligent Systems*, vol. 679, no. 9, pp. 1–12, 2021.
- [12] A. Eke, P. Hermann, L. Kocsis, and L. R. Kozak, "Fractal characterization of complexity in temporal physiological signals," *Physiol. Meas.*, vol. 23, pp. R1–R38, 2002.
- [13] J. B. Gao, J. Hu, and W. W. Tung, "Facilitating joint chaos and fractal analysis of bio-signals through nonlinear adaptive filtering," *PLoS One*, vol. 6, no. 9, pp. 324–331, 2011.
- [14] A. Chhabra and R. V. Jensen, "Direct determination of the $f(\alpha)$ singularity spectrum," *Phys. Rev. Lett.*, vol. 62, no. 12, pp. 1327–1330, 1989.
- [15] J. Y. Sun, J. Lang, Ch. Q. Miao, N. Yang and Sh. Q. Wang, "A digital watermarking algorithm based on hyperchaos and discrete fractional Fourier transform," in *2012 5th Int. Congress on Image and Signal Processing (CISP)*, Chongqing, China, pp. 552–556, 2012.
- [16] Y. Sun, H. X. Zhao, J. Liang, X. J. Ma, "Eigenvalue-based entropy in directed complex networks," *PLoS One*, vol. 16, no. 1, pp. 0–e0251993, 2021.
- [17] Y. Ren, Y. Leng, J. Qi, S. K. Pradip and J. Wang, "Multiple cloud storage mechanism based on blockchain in smart homes," *Future Generation Computer Systems*, vol. 115, no. 2, pp. 304–313, 2021.
- [18] Y. Ren, F. Zhu, K. S. Pradip, T. Wang and J. Wang *et al.*, "Data query mechanism based on hash computing power of blockchain in internet of things," *Sensors*, vol. 20, no. 1, pp. 1–22, 2020.
- [19] W. Sun, G. Z. Dai, X. R. Zhang, X. Z. He and X. Chen, "TBE-net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 0, no. 0, pp. 1–13, 2021. <https://doi.org/10.1109/TITS.2021.3130403>.
- [20] W. Sun, L. Dai, X. R. Zhang, P. S. Chang and X. Z. He, "RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring," *Applied Intelligence*, vol. 0, no. 0, pp. 1–16, 2021. <https://doi.org/10.1007/s10489-021-02893-3>.