Thesis for the Degree of Master

Location tracking technique for Regional ENF Classification Using ARIMA

by

Kim Seo Hyun

Department of Cyber Defense

Graduate School

Korea University

August, 2021

尹 之 源 敎 授 指 導

碩 士 學 位 論 文

Location tracking technique for Regional ENF Classification Using ARIMA

이 論文을 工學 碩士學位 論文으로 提出함.

2021 年 05 月 21 日

高 麗 大 學 校 大 學 院

사 이 버 國 防 學 科

金 書 賢 (印)

김서현의 工學 碩士學位論文

審査를 完了함

2021 年 5月 21日

위원장　　윤 지 원　　　　　印

위　원　　김 휘 강　　　　　印

위　원　　이 상 근　　　　　印

Location tracking technique for Regional ENF Classification Using ARIMA

Seohyun Kim

Department of Cyber Defense

Graduate School

Korea University

August, 2021

Abstract

Electrical Network Frequency (ENF) signal, one of digital forensics technologies, has emerged recently. This signal is useful for finding geographical locations through data training. It is known that the signal uses interpolation based on trained data to find the locations. In this paper, I aim to propose a new location tracking method which does not require the existing interpolation method when trying to find the grids where signal occurs. To do this, I first collect power grid frequency signals from the streaming videos of the online multimedia services before extracting ENF signals from them using the secondary interpolation FFT (QIFFT). Then, I compare coefficients from three different regions – West, East, and Hawaii in the US – by estimating ARIMA (Autoregressive Integrated Moving Average) models. Consequently, I propose the new method to track locations of the streaming videos.

Contents

List of Tables

List of Figures

1 Introduction

With the Fourth Industrial Revolution, new changes are taking place, and the technology sector is rapidly developing. As technology has advanced, the use of computers and the Internet in our lives has become an essential part of society. As the entire economic society is made up of networks and connectivity expands, various individuals, societies, and companies are using online. Furthermore, the industry is creating value using new wireless technologies, and the supply chain relationship is changing. Companies are participating in new connected ecosystems rather than independent operating methods, triggering fundamental changes in the industry, and the emergence and introduction of new technologies are expected to accelerate. In other words, if ongoing devices, networks, and services are hyperconnected, interdependent, and sophisticated shared infrastructure is realized, it will become an era of ubiquitous connectivity across the economy and society.

However, as technology advances like this, advanced illegal activities using these technologies online are also increasing. Starting with individual users, global companies, and even government agencies are being damaged by cyber attacks. In particular, the representative types of information area functions include information leakage and system destruction due to illegal intrusion of information systems, production and distribution of computer viruses, privacy infringement through misuse and abuse of personal information, and spam mail reception. Cyber threats and side effects such as hacking and viruses can lead to crises in our society, information society, and countries, which are absolutely dependent on information and communication infrastructure.

In particular, digital forensics is the process of collecting, archiving, analyzing, and reporting data through logical and standardized procedures and methods that can be used as legal evidence from accessible digital data source repositories. That is, recently, digital audio or video evidence has been used to detect and track illegal activities online. However, since these are also audio/video evidence collected online, there is a problem that tampering or contamination with existing data can occur, such as easily edited arbitrarily by someone. To solve this problem situation more clearly, the use of timestamps to prove that the data used as evidence has not been tampered with is essential. In this regard, the electrical network frequency (ENF) standard is now used as a timestamp for evidence and is emerging as a means of assessing the integrity of digital audio/video evidence and forensics.

In the digital evidence required to verify and verify illegal activities, location information is meaningful data from a digital forensics perspective, which can be used as an important clue to perform circumstantial evidence and proof of absence because it can infer users' travel paths and destinations. In particular, when proof of location information is needed, I usually use location tracking techniques for the device itself, which includes navigation devices. However, assuming that no device is given to collect location information, there is difficulty in obtaining the necessary location data. For example, if video data that is actually provided online in a streaming state is given as evidence, there may be difficulties in finding a specific location value in the service. In this sense, information about where the current video is played can also be important evidence and element in the forensic process. However, with the previously proposed forensic methodologies, there were limitations to how to collect the following data. To address these limitations, a 'positioning method within streaming images' using ENF signals is proposed in the paper. Electric Network Frequency (ENF) forensics analysis is a new method for multimedia authentication tasks such as recording time estimation, timestamp verification, and clip insertion/delete forgery detection, and research has been ongoing to find where the signal originated from. However, ENF signals have limitations in that they are difficult to be collected consistently, and interpolation techniques are commonly used for signals that are not collected to solve these problems. However, interpolation techniques are also inaccurate and typically collect ENF signals from the power grid, resulting in such problems, which require more fundamental problem solving. Accordingly, this paper aims to simply find the originating region based on the ENF signals collected within the online multimedia rather than on the power grid. And machine learning techniques can be used to track specific locations without complicated techniques.

In this paper, I collect data by crawling an online multimedia service that is playing streaming videos. In addition, I reliably extract ENF data with time series features based on speech and location information within the service. I then trained this obtained ENF data using the Automatic Integrated Moving Average (ARIMA) model. Data learned through ARIMA will have a constant number of coefficients. And the coefficients have significant feature values for each region. In other words, this paper introduces a technique that allows these coefficients to track locations through streaming videos played on online multimedia.

The techniques presented in the paper propose indicators that enable ENF signals to track the location of digital audio/video evidence. Furthermore, it is significant in that it is simply possible to determine where the video is played through streaming videos without the interpolation process commonly used for position prediction using ENF signals. This has the advantage of being able to track locations more simply and efficiently than conventional location tracking techniques. In addition, unlike common location tracking techniques, it has great significance in that it can be found without applications or devices that have the ability to track locations.

The composition of this paper is as follows. Following the introduction of Chapter 1, Chapter 2 introduces traditional forensic techniques. It also introduces the concept of enf and introduces prior studies on ENF, referring to the characteristics of ENF forensic techniques that differ from existing forensics techniques. And furthermore, I illustrate how this paper has a different approach from existing techniques. Chapter 3 describes the process of collecting and refining data for use in the paper, and outlines the ARIMA techniques used intensively in the paper. Furthermore, while practically learning the data, this paper extracts specific values that allow us to track the local locations needed to ultimately proceed with the prediction. In the next chapter, I proceeded to determine what significant values the results of the experiment had. To this end, I proceed with classification learning using two machine learning techniques. Finally, Chapter 6 explains what the paper ultimately implies and what significance it has. Furthermore, I discuss the limitations of the research covered in the paper and the future direction of resolution and progress.

2 Background

In the background part, I generally introduce existing digital forensics methodologies. It also compares this to digital forensics using ENF. Furthermore, I elaborate on the process of collecting and extracting real-world data to be used in the experimental part of this paper.

With the development of IT technology, digital and digital technologies have become an essential tool for significant impact on society and individuals worldwide. People are living online communicating, sharing information with others. And the quality of life and convenience of people increased that much. However, this technological revolution in information and communication does not necessarily generate only positive aspects.

Now that digital is an inseparable factor from life in daily life, there are many sophisticated digital crime behaviors using advanced technology. Previously, illegal activities such as fraud, sexual harassment, and intimidation are occurring more widely in cyberspace, and various illegal and criminal activities related to computers such as hacking, music and illegal downloading, and theft are also increasing significantly.

And when such online criminal activity occurs, evidence to punish it is also collected through digital cameras and mobile devices. e.g. Closed-circuit television (CCTV) Among the digital evidence, there are two main areas that need to be noted in this paper. First of all, I pay attention to the video evidence collected from among the evidence. Simple and systematic forensic techniques are essential to proceed with evidence analysis based on collected video data. While such video-based forensic techniques vary in class, they typically use deep learning-based object detection and tracking algorithms or methods to proceed by replacing their images with image data.

Secondly, location tracking data. When the collected digital evidence requires a value for location information, it usually uses location techniques for the device itself that contains navigation devices. Location information is extracted using data from the device itself, and various forensic techniques such as TomTology or Oxygen Forensics Suite are being used for this purpose.

These digital forensics technologies simply perform analysis of the video data itself or use location tracking methods based on the movement of the device. These studies can be extremely significant methodologies when evidence of images in the video is needed, or simply evidence of the movement of the device itself or the person with the device. However, it is difficult to obtain the necessary location data as evidence, given only video data that is actually streamed online without being given the device to collect location information. In this sense, information about where the current video is played can also be important evidence and element in the forensic process. However, there are limitations to collecting the following data through the previously proposed forensic methodologies.

In particular, forensic techniques that track locations lack adequate forensic procedures, have too many different types of navigation devices, uncertain forms of evidence, and limited information available while using non-invasive methods. For this reason, research on location data, which can be used as important digital forensic evidence, has not yet been activated compared to other digital evidence.

To address these limitations, the paper proposes a position tracking method within streaming images using ENF signals.

ENF is the supply frequency within the power grid. In general, ENF has a frequency value supplied from the power grid. ENF is caused by the difference between the power produced and the power consumed by people, which also causes the ENF frequency to change continuously. That is, areas that share power grids have similar frequencies. This means that each region has a different normal value. Thus, most of the Americas now have a typical ENF value of 60 Hz and 50 Hz in Asia and Europe. This means that regions that share such power grids can use ENF as a digital forensics technique, given that they have similar ENF values. Using ENF standards as a means of verifying forensics provides information such as where or when the evidence you want to verify was edited. Based on these characteristics of ENF, this paper discusses positioning forensic techniques using ENF signals.

3 Related Work

Electric Network Frequency (ENF) variational forensics analysis is a new method for multimedia authentication tasks such as recording time estimation, timestamp verification, and clip insertion/delet forgery detection, and research has been ongoing to find where the signal originated from.

Developing forensic tools to authenticate multimedia recording using environmental signatures known as electrical network frequency (ENF) signals arising from power networks is an active area of research.

- ENF signals provide a variety of information for legal audio/video analysis: The integrity of digital audio/video can be assessed through ENF signals. There are three main methods used. First of all, the 'time/frequency domain' spectrograms method uses spectrograms calculations and has significant advantages in speed and implementation over other methods. It is also useful for verifying the date and time when video/audio has progressed through the timestamp of evidence. Next, the 'frequency domain' method calculates the FFT over a short period of time, and extracts a peak value of 50 Hz. This is also a way to identify the date or time when the evidence is recorded. Finally, 'time domain' analysis uses zero-cross measurements. Unlike the above frequency domain method, it is characterized by being applicable to only one ENF element.
- ENF signals can be used as timestamps in the forensic process of video data: I identify changes in the image caused by fluorescent lamps in videos taken indoors and separate the video into segments of regular intervals to use them. In addition, the ENF signal from those parts is compared to ensure that the ENF value of a particular part is

not different from that of others. If a particular part of the video is arbitrarily changed or inserted, the ENF signal for that part is different from the value of the other segments, so this verification can confirm that the image has been modified. However, these analyses are not valid to be used as a method to verify that certain regions have been deleted, as they are to divide video images into segments and compare them to each other.

– ENF have a specific value for each position: When the power grids are independent of each other, they have different values when collecting and comparing specific ENF signals from regions. Independent power grids simultaneously have similar ENF signals. Therefore, ENF values have very similar values in the same power grid, and it is difficult to grasp the meaningful relationality of ENF values in regions in different power grids. That is, in this sense, ENF signals can be referred to as time series data. And if I check the data from a time series data perspective such as ENF, the time series data are typically based on correlations between continuous data when processed. Furthermore, the variation of ENF signals at the grid level can be distinguishable between records performed on different grids, as they generally differ simultaneously between independently operated grids. In other words, it is an explanation that proves that ENF signals have unique characteristics of a particular region and typically construct protocols to verify a particular location. The technique used here is based on the semi-planar crossing method. In other locations, changes or variations in ENF can be explained by signal processing algorithms. A method of proving this variability is to use similarity measures proportional to the distance between locations. And under certain conditions, the proposed protocol is shown to provide very high estimation accuracy.

– ENF Signal is illustrated through various estimation methods: ENF is computed using Fourier transform-based frequency estimation methods. ENF signals are extracted from records by filtering operations and then extracted via instantaneous frequency estimation. For timestamp authentication and verification, similarity between multimedia and ENF signals extracted from the power database can be measured at that time. Here, I use a regularized cross-correlation coefficient (NCC). If the NCC value of a particular part is high, it can be determined that that part represents the time at which the recording occurred. Furthermore, there are high-resolution frequency estimation methods such as MUSIC and ESPRIT. This has the advantage of providing better immediate frequency estimation of ENF signals for short segments compared to Fourier transform-based methods.

In conclusion, I can see that the electrical network frequency criterion is very useful as a digital forensics tool. In summary, verification of integrity is a necessary step when analyzing digital audio and video records. To this end, ENF signals can identify when a particular event occurred. It is also a tool that can be used to identify specific areas on a 50 or 60 Hz basis. This is achieved using a reference frequency database where ENF signals are recorded. This ENF criterion is very significant when used in conjunction with existing forensic audio and video technologies. In addition, synergy can occur when used in conjunction with other methods, such as IT investigation using various digital technologies, rather than just ENF signals as a forensic tool. In other words, forensic digital forensics using ENF signals should be able to sufficiently improve standardized analytical protocols and integrate naturally with existing methods. In conclusion, the ENF criterion can be an important breakthrough in finding forensic technology.

In general, the step of interpolating data is essential to proceed with forensic verification using ENF standards. Interpolation is a method of estimating ENF in the unknown domain because the ENF value cannot be measured at all points in the power grid, and the interpolation technique is based on a statistical approach. When interpolation techniques are commonly used to address the limitations of ENFs, inverse-weighted (IDW) interpolation techniques are used to estimate ENF values in unknown domains. IDW can also be used as an outline estimation algorithm, and the point estimation value in the IDW algorithm is defined as the sum of the weighted reliability in inverse proportion to the distance. This is described in detail as follows:

Reverse-weighted (IDW) interpolation: When ENF uses inverse distance-weighted (IDW) techniques to restore ENF signals in an unsampled region, one of the techniques for interpolating contours for IDW is to interpolate empty data based on the reliability of weights in inverse proportion to distance. Direct experiments have shown that the squared distance is inversely proportional to the squared distance, which allows us to estimate the ENF value of regions where ENF is not extracted based on a specific confidence interval. This interpolation allows the extraction of existing ENF values of leaked packets using different ENF values even if the ENF breaks in the middle, and is a good technique for significantly increasing reliability by allowing the full name value to be stored on a specific power grid.

The formula(1) for IDW is as follows.

$$u(x) = \frac{\sum_{k=0}^{N} W_k(x) u_k}{\sum_{k=0}^{N} W_k(x)}. \tag{1}$$

IDW is a good way to solve the limitations of ENF, but it is difficult to fully restore ENF signals in areas where this interpolation method has not collected.

4 Data

This part describes what real data will be used in the experiments of this paper and explains how it was collected.

4.1 Data Collection

There are many ways to collect ENF signals, but usually ENF data is collected from the power grid where the FDR device is installed. Studies show that the number of FDRs deployed across the U.S. is 129 in east, 45 in west, and 7 in Texas. He also explores the FNET/GridEye website to actually collect ENF data. The FNET/GridEye website usually updates ENF data every four seconds, but delays can occur depending on the state of the instrument or the external environment. For this reason, it is necessary to ensure that the FDR has failed or stopped unexpectedly, or that the FDR is not too old. In addition to FDR, it is necessary to continuously check for other defects, such as the associated network connection status, and to ensure that no data loss has occurred. In addition, data collection will result in a collection interval or a separate anomaly, and additional linear interpolation should be applied.

However, unlike the previous methods, this paper collected the data needed for experiments within an online multimedia service with networks within the U.S. power grid. The online multimedia service used to collect has the following characteristics:

- Provide location information (hardness)
- Use alternating current for video recording.
- Provide continuous and real-time audio data

This means downloading audio files from the specified online service that contains ENF signals. It retrieves html files to analyze geolocation information provided by online multimedia services. The default protocol for video streaming is m3u8. The m3u8 protocol can continuously download files from the media server. The module continuously downloads streaming videos to collect voice data. The latitude and longitude of the webcam recording are in the html file on the website. Then, I create a program which both finds and collects values that represent latitude in html. This program runs as a Python wrapper to download audio files in real time and store the collected audio. To parse this data, use the Beautifulsoup4 library, which makes it easier to parse data from HTML files. And I checked how the web service is proceeding before the crawl.

And as a result of checking, the web page works dynamically, so I wrote the code according to the dynamic environment. Calling a page defined by a URL is called a static page, and using the method of importing data from the top of a web page is called a dynamic page, as shown in the following map. Dynamic environments generally have the disadvantage of being slower than static environments, but they have the advantage of being able to collect continuously and not having separate limits to the collection area. In other words, dynamic pages change the source code according to the behavior of the user as they are accessed, so they need to be approached in a different way. Pandas saved the crawled data as a csv file and created an object for the Web Driver.

After saving the latitude, longitude, and m3u8 files to the streaming video, the audio was downloaded via it and resampled using the ffmpeg tool. ffmpeg can download audio files via the m3u8 protocol. This means that the ffmpeg tool is used to download and resample audio from the m3u8 address. Because the default sampling rate for audio files is too high, it requires a large amount of memory to store the original audio file. Therefore, the audio file is downsampled up to 1,000 Hz, which does not compromise the existing ENF signal in the audio. Finally, the acquisition module stores downsampled audio files, latitude, and time information along with hardness of 1 kHz.

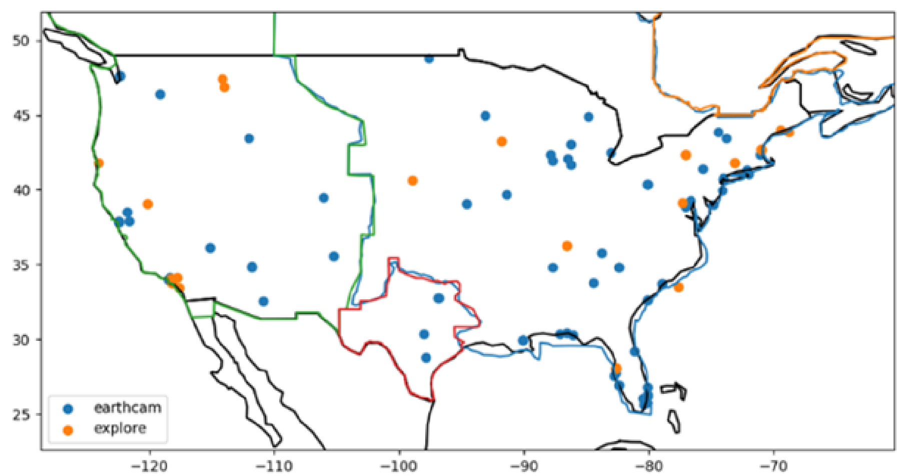

Fig. 1. Collected data from online multimedia services

As such, the streaming service collected the location and voice data for the experiment. The collected data is shown in the figure 1. Next, I collect ENF signals from this collected data. In fact, the collected data was confirmed to be in the eastern United States and Hawaii. In addition to the U.S., data from various regions such as Europe and Africa were collected, but it was confirmed that it was not significant enough to conduct experiments and that a large number of data were collected. If you think about the reason, you might think that the power grid was not clearly built like the U.S. power grid, or that the video provided by streaming services usually occurs in the East and West of the U.S. and Hawaii regions.

4.2 ENF Extraction

The data I collected in the above way is gastric and longitude data based on Google that can identify voice/image data and location. Since the practical data to be used in the experiment is an ENF signal, the process of extracting the ENF signal from the voice data in the image is necessary among these collected data. And this extraction process typically uses modules, which require the following functions:

- Extraction algorithm: I estimate algorithms of ENF signals (e.g., QIFFT), which require a variety of preprocessing techniques.
- Maximum likelihood estimation (MLE): The MLE enhances the ENF signal with a harmonic signal (120 Hz). This step is very important for partially distorted signals. Even if the ENF value of the underlying frequency is distorted, signals of other frequencies may be of higher quality.
- Filter: The filter removes abnormal ENF signals that are not correctly extracted from the Threshold Dependent Medium Filter (TDMF). ENF signals can be distorted due to background noise. Once the extracted ENF signal passes through the filter, it can be stored in the database.
- Database: The extracted and filtered ENF signals are stored in the MySQL database. The format of the stored data is (information about latitude, longitude, time, and enf source).

Specifically, the specific part of this paper is the process of filtering unusual signals that can occur when background noise is approximately 60 Hz (base) and harmonies. After acquisition, the ENF signal extraction phase extracts the ENF signal from the file using secondary interpolation FFT (QIFFT). In addition, STFT (Short-Time Fourier Transform) is used to efficiently analyze information about frequency changes over time in the signal. These STFTs are also calculated with frequency (ENF signal) resolution as 'FFT number/sample frequency'. For example, if the sampling rate is 1000 Hz and the number of FFTs is 1024, the FFT resolution is approximately 1 Hz (=1000/1900). By schematicizing it, it can be expressed as shown Figure 2.

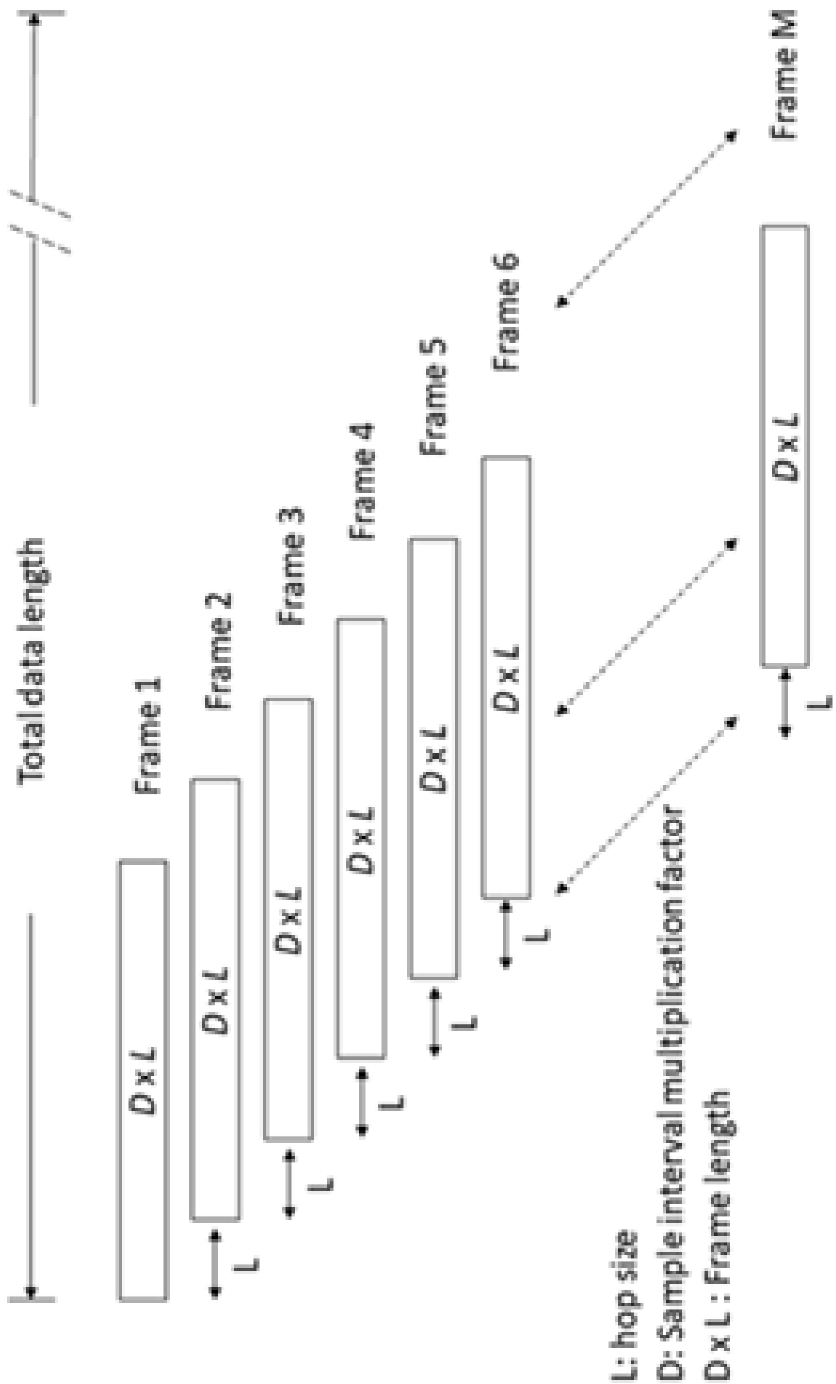

Fig. 2. FFT frames of the Short-Time Fourier Transform (STFT)

However, since the ENF signal has a low standard deviation, additional steps are required to increase resolution using QIFFT. QIFFT is an interpolation method that uses a quadratic polynomial, where three data are given, a polynomial passing through this point, and where the polynomial can be used to obtain interpolation values for a given point. That is, the STFT result is applied as QIFFT, so the ENF signal is estimated to be one target frequency v. The frequency estimated by QIFFT can be calculated as shown in (2) and Figure 3.

$$p = \frac{1}{2}\frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \tag{2}$$

$$(\alpha = 20log_{10(X_m(k_{\beta-1}))}, \beta = 20log_{10(X_m(k_{\beta}))}, \gamma = 20log_{10(X_m(k_{\beta+1}))})$$

As a result, I collected Google-based latitude location information and voice data within online multimedia. And based on the data collected like this, I extracted ENF signals using the QIFFT algorithm.

The ENF signals extracted from voice data are shown in Figure 4. ENF signals for each of these extracted regions are shown in the following figure 5,6,7.

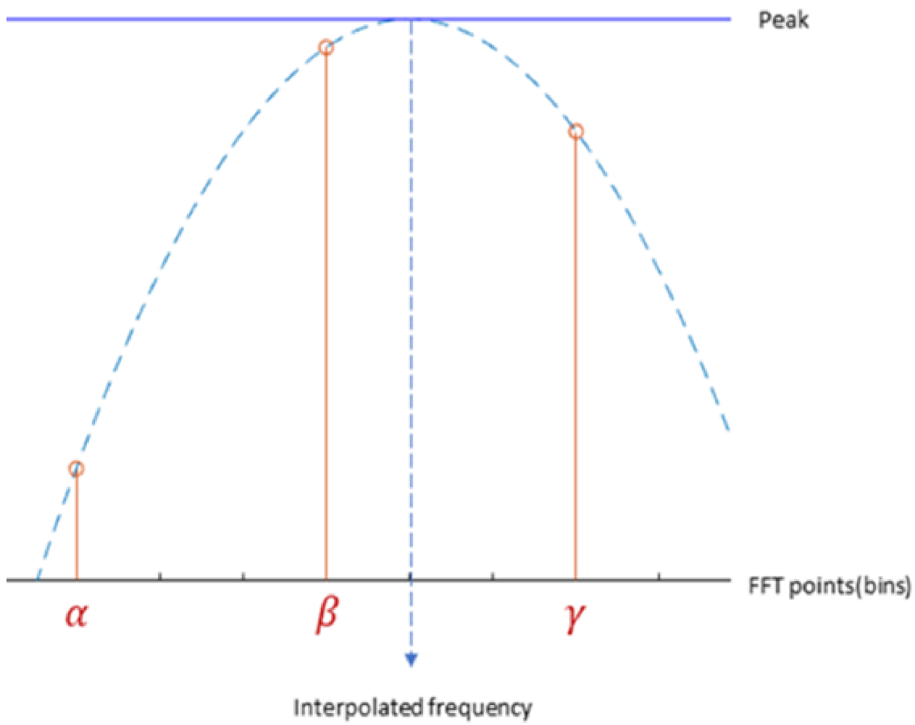

Fig. 3. QIFFT (Quadratic Interpolated Fourier)

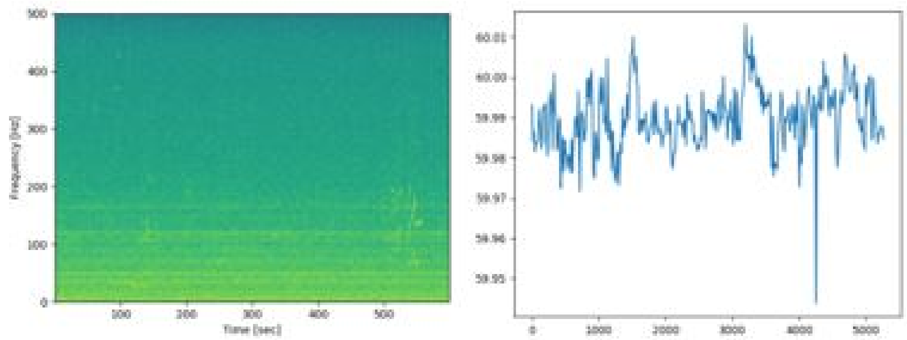

Fig. 4. Spectrogram and ENF signals of audio file downloaded from online multimedia service

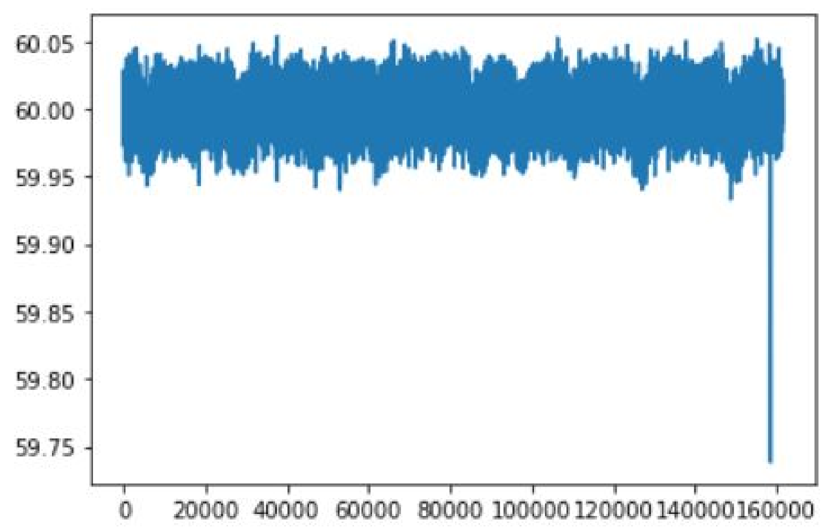

Fig. 5. ENF data (East)

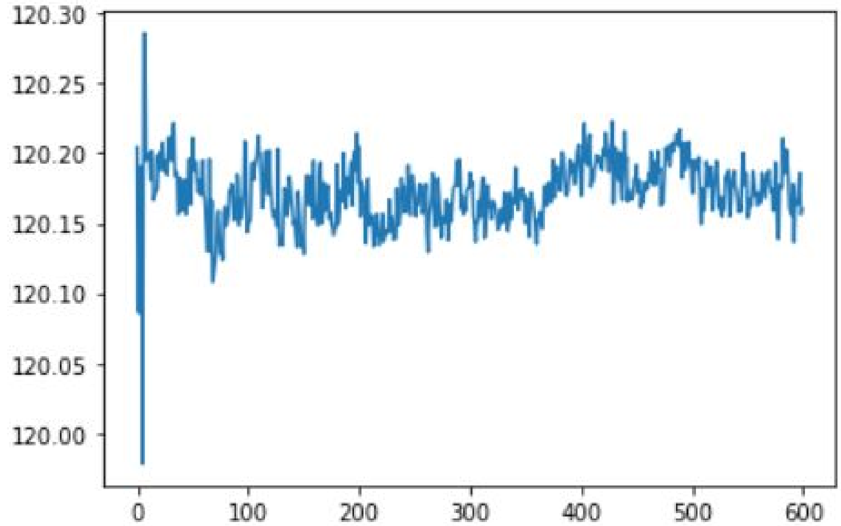

Fig. 6. ENF data (Hawaii)

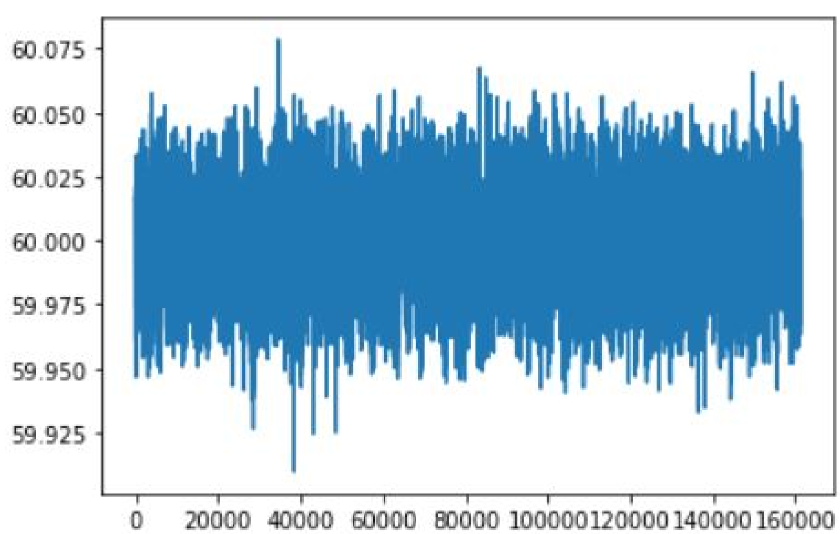

Fig. 7. ENF data (West)

5 Technique

Before proceeding with the experiment with the data collected like this, the main technique used in this paper is ARIMA technique. ARIMA is a frequently used technique for time series data. So before I explain ARIMA technique, I will explain time series data.

5.1 Time-Series Data

Time series data is a set of sequentially determined data sets collected over a constant period of time. The characteristics of the time series data are that they are ordered over time, and consecutive observations correlate with each other. In other words, the purpose of analysis of time series data is to discover the lawfulness of the time series, model it, and forecast future values through the estimated model.

5.2 Stationarity

First of all, data must have characteristic properties before analyzing time series data. The most basic property of time series data is stationarity. And for this stationarity to be established, it requires three major properties: the mean converges to a constant value, has finite variance that does not change over time, and the correlation coefficient decreases as the delay length increases.

5.3 Autoregressive Integrated Moving Average (ARIMA) Model

In this paper, an autoregressive integrated moving average (ARIMA) model, one of the time series analysis techniques, is used. AR = autoregressive; I = integrated; and MA = moving average.

The ARIMA model is represented by a formula as follows.

$$y_t^{'} = c + \phi_1 y_{t-1}^{'} + \cdots + \phi_p y_{t-p}^{'} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \qquad (3)$$

where the stationary dependent variable, yhat t is a function of a constant, c is the constant,are the parameters of the autoregressive part up to p, thetaq are the parameters of the moving average part up to q, and varepsilon are error terms. To describe the ARIMA model, it is necessary to identify how this model is constructed and explain the characteristics of each element.

First of all, it is necessary to explain the AR model. The p values that can be specified in the AR model indicate the order of the autoregressive process. This reflects the characteristics of time series data, which means that data accumulated in the past affects subsequent data changes. That is, when present is a dependent variable on a time series, the value of time series data that occurred at a previous point in time is taken as an independent variable. In other words, if you check the AR model based on this concept, the p value that can be specified in AR indicates the delay value of the dependent variable in the model. At this point, the AR(p) model can be defined as stationay only if the root of the polynomial equation is greater than the absolute value of 1.

And in general, it is essential to check the Cor value for the autocorrelation function and time series values to determine the exact p value of the model depending on the data. So, for example, if the autocorrelation function AR has a p of 1, the Cor value for the autocorrelation function and the time series value can be exponentially reduced if the current value reflects a large number of previous values. In addition, the Partial Autocorrelation Function (PACF) includes the process steps of the AR model displaying the estimated coefficients from the OLS estimates. If the observations are generated by the AR(p) process, you can see that the theoretical partial autocorrelation results in zero for parallax exceeding p.

MA model, which is the component of ARIMA model, is based on time series data characteristics just like AR model. The tendency of the mean value of the data itself to change over time is called Moving Average. This means that the moving mean is constructed into a time series model is called the MA model, which takes advantage of the historical value of the forecast variable in regression. The difference between the autoregressive model and the moving average model is that they are calculated using historical prediction errors in models that look like regression. In other words, the error before that is the central part, and for the previous value, it is not important in the MA model.

In other words, the value specified by the MA model, q, represents the average value of white noise. That is, if in a time series q has a constant value as the mean of white noise, the MA model will have normality. And in this regard, the unique characteristic of MA models is reversibility. MA models, like AR models, use ACF and PACF to check time series sequences. All MA processes can be represented as AR processes with geometrically decreasing coefficients, so the PACF of the MA processes will appear to be decreasing.

The model that combines the AR and MA models described earlier is ARMA model. ARMA deals with the order value of the autoregressive polynomial with p and the order value of the moving average polynomial with q. The properties for the stationary of the data cover only the ARMA(p,q) part of the AR(p). That is, the proot of polynomial equation

(z) = 0 must be outside the unit circle. In the same vein, the properties of reversibility for ARMA (p, q) models relate only to the MA(q) portion of the specification, and the root of the polynomial must also be outside the unit circle.

In turn, I explain the 'I' part of the ARIMA model. ARMA models can only be created with a still time series. This means that the mean, variance, and covariance of the time series are constant for all times. The ARIMA model is not only related but also considered public figures. There is a linear relationship between two variables X-Y that if the correlation is greater than zero, the Y value has a large value when X is greater than zero, and Y has a small value when X is greater than zero when the correlation is greater than X-Y. Also, if the coordination is greater than 0 between the two variables X-Y, the Y value increases if the value of X increases above the previous value, and if the coordination is less than 0 between the two variables X-Y, the Y value decreases.

Put differently, if correlation is less than zero and Cointgration is greater than zero, then it will have a value that increases faster and faster. Therefore, ARIMA models go through the process of checking the differences in order to maintain their stickiness. In other words, you need to identify differences and eliminate trends in one-data. In other words, while the ARMA model uses historical data, the ARIMA model reflects the "momentom" that historical data had, which only responds to its own (normal data) trend and does not consider white noise.

In other words, the variation of the dependent variable is associated with the value of the constant, and in the ARIMA model, p and q represent the order of the autoregressive model and the moving average model, respectively. p refers to a value that has a linear relationship with the previous value, and q is the weight value for the previous deviation.

The values p and q depend on the characteristics of the data, and data analysis can determine whether the data are suitable for AR, MA, or ARIMA. That is, ARIMA, which contains the conditions of stickiness and reversibility to the data, is a model that reflects the autocorrelation that the previous value affects the subsequent value of the random variable and the moving average phenomenon in which the previous event affects subsequent results.

To identify the characteristics of the previously collected time series data, I analyze ENF signals from three regions of the United States, East-West and Hawaii. I assume that the prediction of values for future periods of time series data, defined by a sequence of observations arranged chronologically, is determined by the influence of the past observations of the variable under study. Therefore, when analyzing time series data, the first step is to induce the data to have a seasonality and a trend that does not exist.

There are many ways to verify the stickiness of time series data. Among them, the statistical test method, 'Augusted Dickkey-Fuller test', can accurately determine the properties of the time series data by checking the extracted p value. The augmented Dicky-Fuller test for each data confirmed that the p value was '0.0' and that the ENF data for all regions were time series data in a stationary state.

6 Experiment

I conducted an experiment to track the ENF signal generated from the power grid and predict where the corresponding image is being played. First of all, ENF data was extracted from voice data collected within online multimedia. And I will learn the extracted data in the ARIMA model described earlier. And the coefficient values of p and q extracted through the ARIMA model are taken. This was carried out in the East/West of the United States and Hawaii respectively, and the process of verifying the characteristics of each coefficient with a machine learning model was carried out.

6.1 Training

It takes several steps to train previously prepared data to ARIMA models. First of all, a data refining process was needed. First of all, my experimental consideration is that when time series data are applied to ARIMA models, only one coefficient is extracted when time series data are applied to the learning model at once. This could lead to the problem of too little data being used for future supervised learning stages. By providing the above training data to the machine learning model through supervised learning, the model finds a rule that can create a target value with a given input value. Then, it use the rule to make predictions for other inputs, and then find the difference from the actual answer. At this time, if this difference is large, the learning is carried out by modifying the rules.

In fact, creating training data is more important than building models in supervised learning. This is because including incorrect inputs and targets in the training data can lead to incorrect models, and too few data can sufficiently train the model. Furthermore, despite the fact that the data was collected over the same period of time, the number and size of ENF signals collected in each region are different. The basic limitation of ENF is that it is technically affected by other environments with the characteristic that the signal is variable, as mentioned earlier. Therefore, interpolation techniques are basically used in the process of refining ENF signals. In the same vein, I were able to confirm that the ENF signals available for training had different sizes.

To address these issues, I introduced a method of grouping collected regional data into groups. Each region's data was randomly divided into 10 groups and trained in the model through ARIMA. In this way, even if the number of ENF signals collected over the same time period varies from region to region, the number of coefficients available for subsequent prediction processes is consistently extracted to 10. In other words, 30 coefficients are extracted for all regions. In addition, I divided it into groups of 100, 200, and 1000 so that certain coefficients were extracted by each size.

Before applying this refined data to an experiment, the characteristics of the previously mentioned time series data should also be applied to the data for future experiments. To do so, the ARIMA model must be determined by identifying the autoregressive and moving average elements of the ARIMA model after passing the normality test of the time series data.

In other words, in order to adequately set p and q values for AR and MA, I theoretically set up and analyze the model by comparing autocorrelation functions (ACF) and partial autocorrelation functions (PACF) to identify the proposed model by Box and Jenkins. As mentioned earlier, the autocorrelation function (PACF) and the autocorrelation function (ACF) can be used to determine the autocorrelation mean model p,q order. The method allows us to classify white noise in the data to find meaningful delay values. This means determining the final ARIMA model through parameter estimation and model diagnosis, which verifies the suitability of the subsequent model, and calculating the statistics based on it.

And based on the characteristics of time series data that historical data will affect the future through these inferred statistics, I infer future predictions on the premise that trends similar to the past will continue in the future. In these parameter estimation steps, estimation for a time series model uses a method similar to the least squares method in regression. Based on the Gauss-Newton idea, I apply a computerized numerical analysis method. If the time series data are characterized by AR, the ACF decreases slowly and the PACF decreases sharply except for the first time lag. Conversely, if MA is characterized, ACF decreases rapidly and PACF decreases slowly. The values at this rapidly decreasing point can be used as parameters (p, q) for each AR and MA model. In addition, the data can be calculated by calculating ACF and PACF by calming down the data, resulting in an appropriate number of differences.

The results of analyzing the actual data used in the experiment can be referenced in the illustration. In fact, there was a difference in the order values of ARIMA models suitable for East-West and Hawaii data. First of all, as a result of analyzing data from eastern and Hawaii, the PACF coefficient decreased smoothly. However, unlike PACF, the ACF graph has drastically decreased from the second delay value. Since the graph determines that only the values of the correlation coefficients are significant at the level before the plunge, it shows that the data in East and Hawaii are suitable for the MA model of ARIMA. In other words, the exact order of the East and Hawaii data is (0,0,2). On the other hand, West data showed that both PACF and ACF graphs saw a sharp decrease in the delay value from the first delay value. That is, western data fits the ARIMA model in the appropriate order (1,0,1). This is shown in Figure 8,9,10.

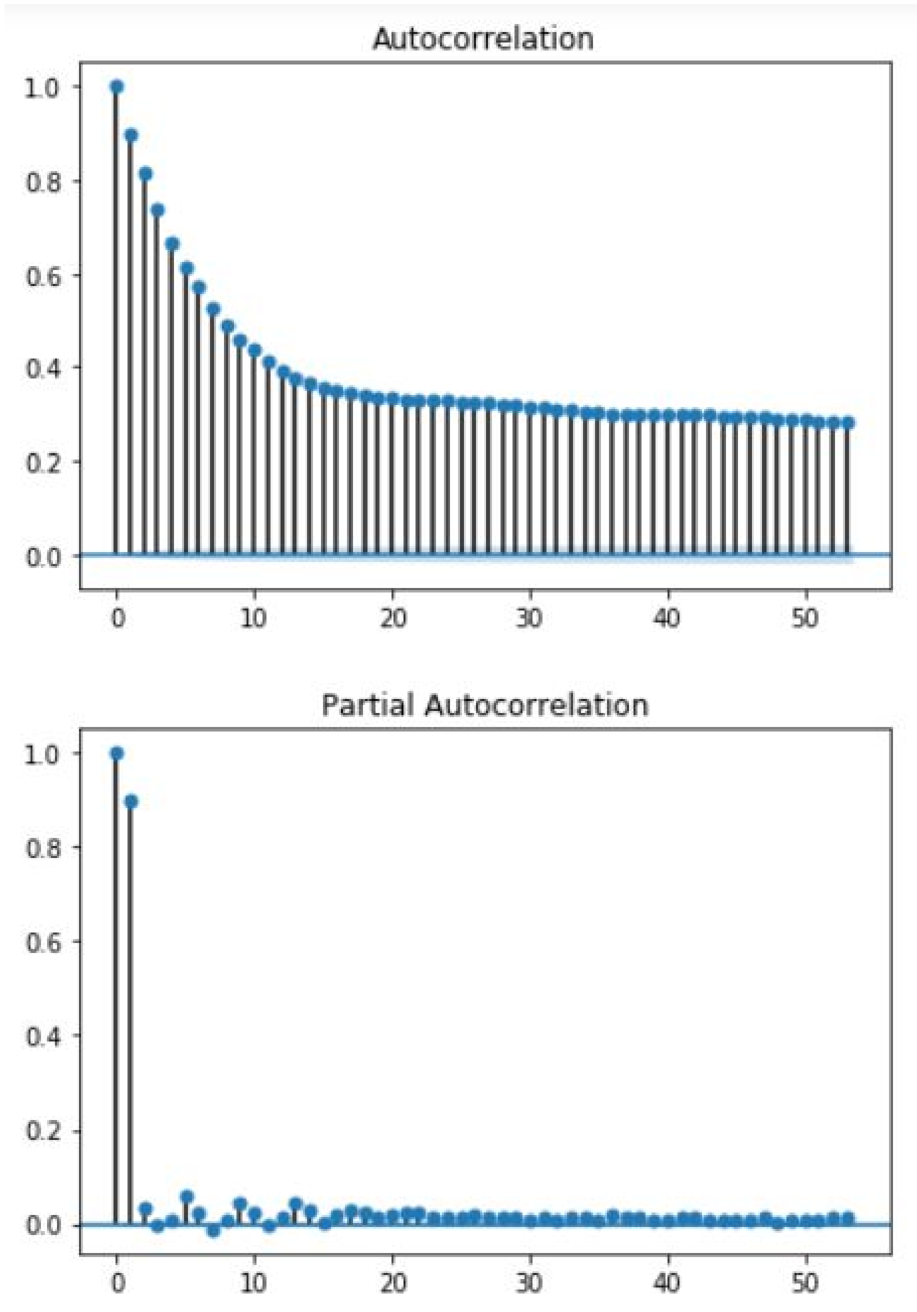

Fig. 8. ACF and PACF (East)

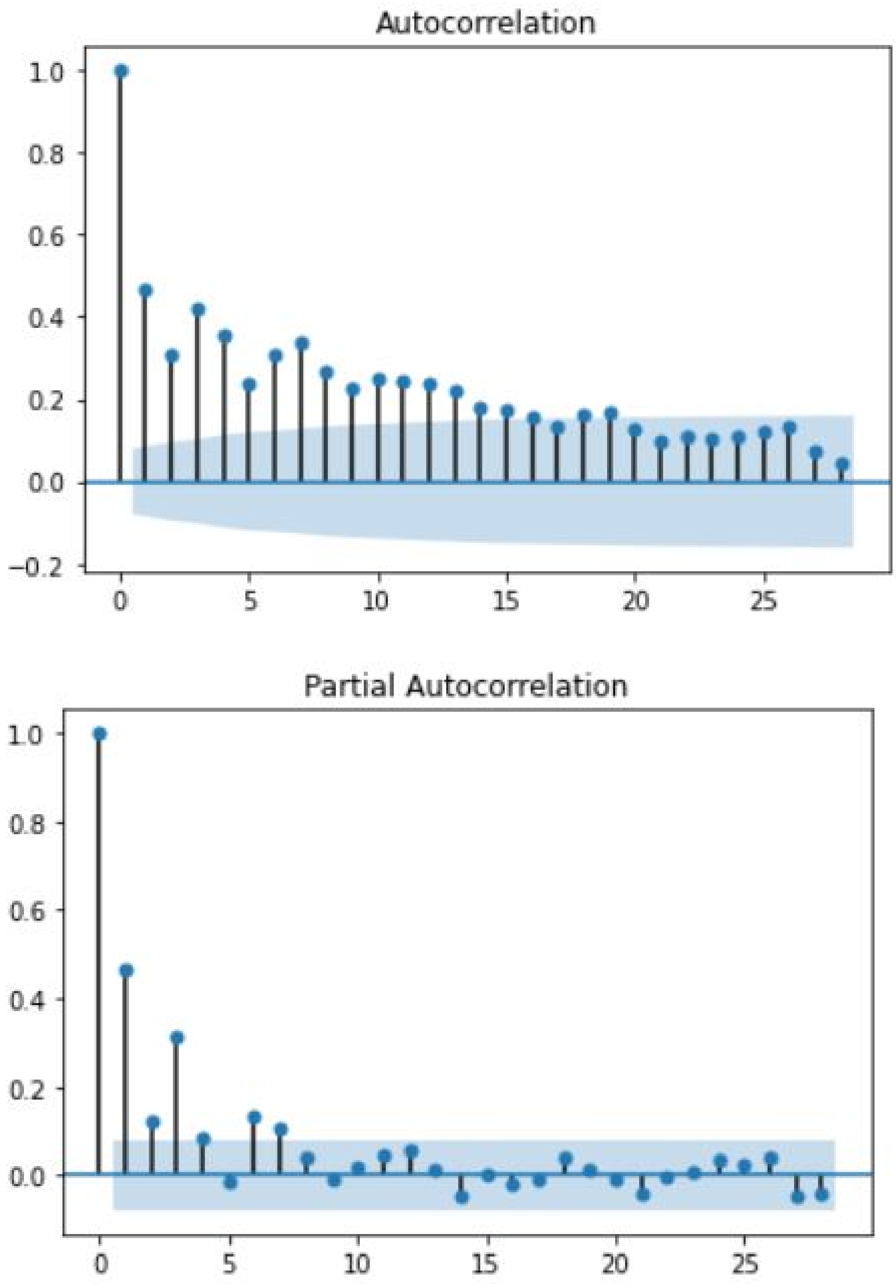

Fig. 9. ACF and PACF (Hawaii)

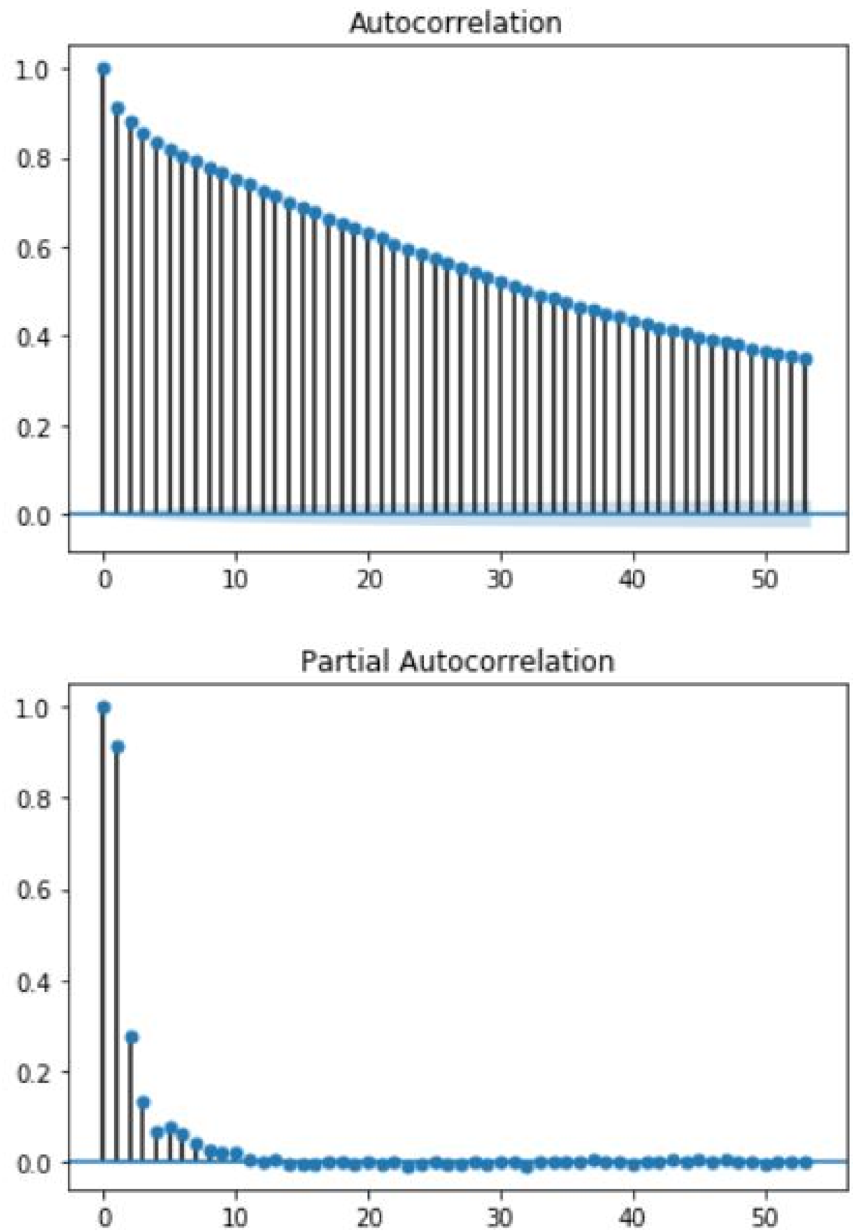

Fig. 10. ACF and PACF (West)

Finally, during the Model Diagnostics phase, I apply the autoregressive moving average model and verify the normality of the remaining residuals. At this time, it is verified whether the residuals are independent of each other, or whether the mean and variance are constant over time. To this end, I test the Ljung-Box statistics that test the autocorrelation of time series data, whether the means and variances are constant, and whether there are any additional missing from the model using autocorrelation functions and partial autocorrelation functions. Subsequently, the optimal model for the time series data is further predicted through a prediction function when an appropriate order of autoregressive shift means is adopted according to information theory (AIC, BIC).

I applied the ARIMA coefficient value that was set through refining data through the previous process and applied it to the model. And the results of this application to the ARIMA model are as follows. For all Eastern/West/Hawaii data, the coefficients for the three regions are shown in Figure 11.

If you look at the figure, you can visually see that the coefficients derived from the ARIMA model have significant characteristic values for each region. In other words, as the displayed results show, I can confirm that the extracted coefficients have important properties that can be classified according to local order values. To clarify this classification, I proceed with it using machine learning techniques.

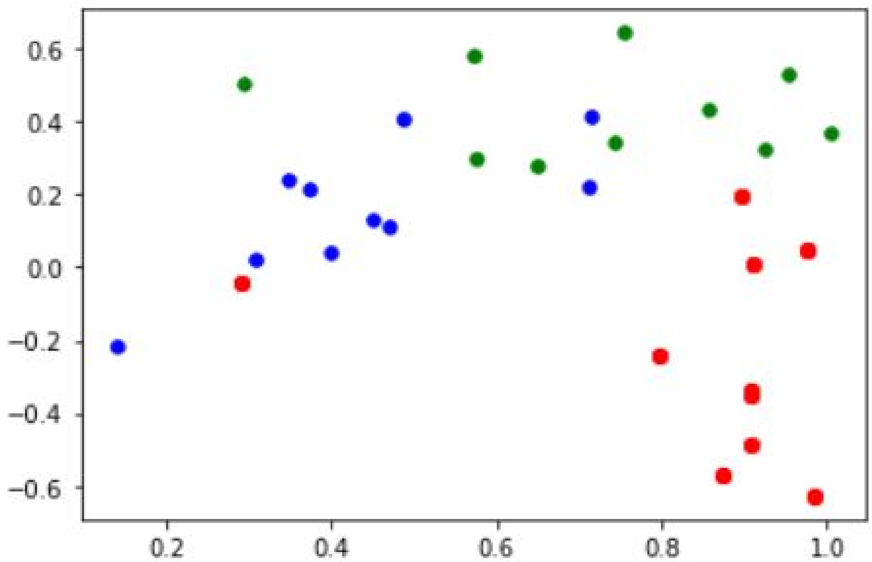

Fig. 11. ARIMA coefficient value of ENF data collected from images within online multimedia (proceed to 10 groups), East (green), Hawaii (blue), West (red)

To verify that the ENF data extracted within the online multimedia service that I actually conducted has an accurate distribution, I further compare it with the distribution of ENF data coefficients collected in different ways. I verify the distribution by applying ENF data collected using FNET to ARIMA models alike, as in methods that model ENFs collected from online multimedia services. As a result, it was also confirmed that each region had a specific coefficient value. This is shown in Figure 12.

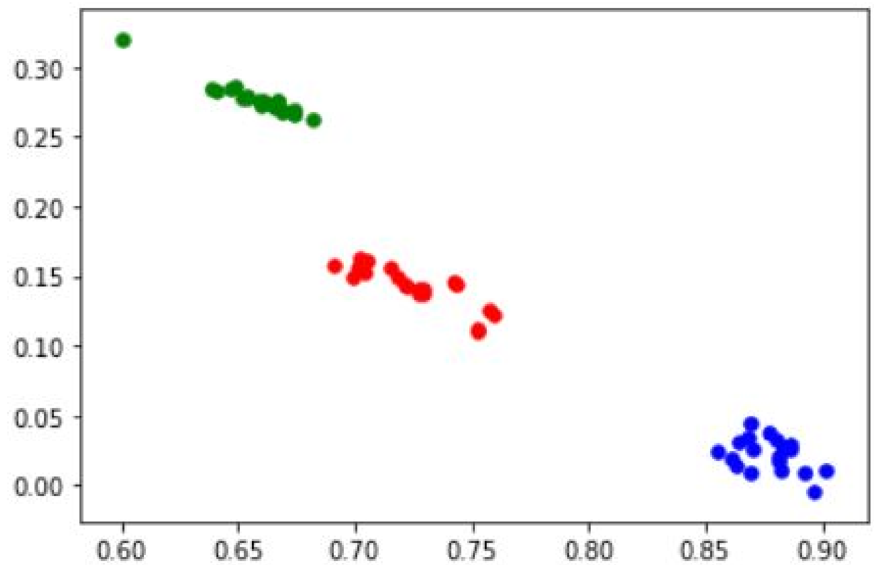

Fig. 12. ARIMA coefficient value of ENF data collected using FNET (proceed to 20 groups), East (green), Hawaii (blue), West (red)

6.2 Classification and Prediction

I put ENF signal data extracted to input values of machine learning classification techniques to understand the unique characteristics of the collected data through experiments. Machine learning classification techniques vary, but in this paper I am a Support Vector Machine (SVM) and a kneighbors classifier. Both of these classifiers are based on supervised learning. To briefly explain supervised and unsupervised learning, the following is true.

First of all, supervised learning, which I deal with in my paper, implies a method of training machine learning models while providing input values and targets. Here comes the very important concept of a loss function. A loss function is a function on which the rules of the model are modified, defined by the loss function = the correct answer – the predicted value. In supervised learning methods, the concept of a loss function may exist because the training data have the correct answer. Instead, fact-guided learning methods are more difficult to create training data than to create models. There are various types of supervised learning, including support vector machine, hidden Markov model, regression, and neural network. Other tasks use features called Predictor Variable to predict the final results, often referred to as regression models.

Unsupervised learning, on the other hand, is a method of training models using targetless training data. For example, an entity might consider a situation in which the entity designates groups according to the customer's propensity to consume. Since I don't know how many groups will be formed or what groups will be formed, I typically cluster them in situations where there are no targets. At this time, it is characterized by the difficulty of evaluating the model's training results because there is no target. The hierarchical clustering algorithm allows you to refine each group into smaller groups. Furthermore, visualization algorithms create 2D or 3D representations that can be schematic when large amounts of unlabeled high-dimensional data are inserted. In unsupervised learning, dimensionality is used to simplify data without losing too much information. For example, a car's mileage is so associated with the model year that two characteristics can be combined into one characteristic that indicates the degree of wear on the car with a dimension reduction algorithm. This is called Feature Extraction. Unsupervised learning is also an essential step in detecting outliers, which automatically removes strange values from datasets before injecting them into the learning algorithm. It is trained with a normal sample, and it is determined whether the new sample is normal or not. Learning association rules finds interesting relationships between properties in large amounts of data. It is used to find that a person who purchased a product tends to buy another product.

In addition, reinforcement learning is similar to supervised learning in that it provides answers, but the difference is that it provides rewards. The goal at this time is to get as much reward as possible, so I learn by performing so many actions in a given environment and repeating the highly rewarded performance.

In this paper, I use supervised learning. Specifically, I employ Support Vector Machine (SVM) and K-Nearest Neighbor.

Support Vector Machine (SVM)

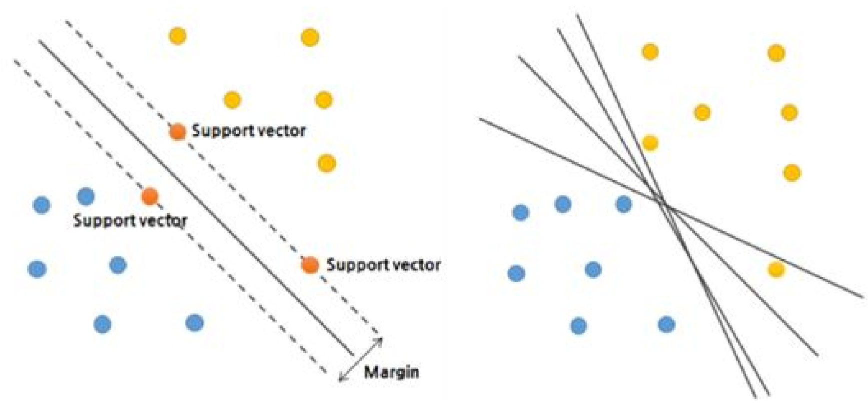

Fig. 13. Support Vector Machine (SVM)

One of the machine learning, Support Vector Machines (SVMs), creates a non-probable binary linear classification model based on a given dataset to determine which category the new data belongs. This is described in detail as follows: A support vector machine consists of a set of superplanes (hyperplanes) or superplanes available for classification or regression analysis. Intuitively, if the superplane has a large difference from the nearest learning data point, the classifier error (English) is small, so good classification requires finding the superplane with the closest distance to the learning data for any classification point. At this point, the data point used to find the superplane is called a support vector. At this point, two parameters must be determined before performing the analysis.

- Cost: Parameters of error tolerance. Adjust the width of the margin.
- Gamma: Parameters related to the kernel, not superplane. Controls the curvature of the decision boundary.

This means classifying the data by changing the linear classifier to a nonlinear structure. A typical case is the classification of data using the kernel method, which is a computational technique used to expand the variable space to accommodate nonlinear boundaries between classes.

In conclusion, when predicting with a support vector machine, the analysis procedure is as follows.

- Get the data you need, preprocess, and generate learning and test data
- Getting parameters for performing support vector machines
- Perform prediction using support vector machine using secured parameters

The advantage of Support Vector Machines (SVMs) is that the prediction accuracy is high. In addition, it is easy to use and has many libraries that support SVMs, and can be applied to areas with low learning data.

The disadvantage is that if multiple classifications exist, each classification must have its own SVM. Also, more classifications consume more resources and take longer.

So, as a result of experimenting with the use of support vector machines in practice, support vector machine learning for each of the three regions has been achieved. And the classification took a long time.

K-Nearest Neighbor

K-Nearest Neighbor (KNN) is one of the supervised learning algorithms. The K-Nearest Neighbor classifier classifies the data by weighting them according to the distance to the neighborhood when the input consists of the k nearest training data in the feature space. That is, it is extremely intuitive, simple, and given some data, I look at the data around it (neighborhood) and then classify it into categories that contain more data.

The characteristic of KNN is that training is not necessary. Other models are trained using code from clf.fit (xtrain, ytrain). It is a method of creating models based on training data and testing them with test data. But KNN doesn't need any training. Just storing training data is all about training. For example, SVM creates a Decision Boundary based on training data, and classifies test data through the Decision Boundary. As you can see in the picture above, KNN classifies new data by looking at the k data around it. Therefore, there is no need for pre-modeling. Real-time predictions are made. It is called Lazy model, meaning that the model is not built separately. Therefore, it is faster than SVM or linear regression.

In KNN, you need to find the distance between the data and the data. There are two ways to find distance.

- Euclidean Distance: This is usually the way to find the distance between a point and a point.
- Manhattan Distance: It means the distance along the x-axis, the y-axis, not the straight line between the point and the point.

In turn, there are two main points to note.

- Normalization: Therefore, when using the K-Nearest Neighbor algorithm, I do Noramlizaion to reflect all the properties evenly. There are several ways to regularize, the two most widely used methods.
  - To fix the minimum value to 0 and the maximum value to 1, and then convert all values to values between 0 and 1
  - To transform the mean and standard deviation into a z-score how far it is from the mean
- Select the number of K: In K-Nearest Neighbor classification, the choice of k counts is important. When k is too small, overfitting occurs and classification accuracy is bound to be significantly lower. This is because it is sensitive to one point nearby. In summary, the K-Nearst Neighbors algorithm results in overfitting when it is not fully considered by other neighboring neighbors. So, if there is one outlier, the label of the nearby point can be determined by the outlier.
  Conversely, underfitting occurs when k is too large. If the k is too large, the classifier does not look at the learning data in detail enough, and accordingly, the classification will not be done well.

KNN algorithms, like support vector machines, have advantages and disadvantages. First of all, the advantage of KNN is that it performs quite well if the number of learning data is sufficient without significantly being affected by the noise of the learning data. Furthermore, considering the aforementioned distance computation and the variance of the data, a fairly robust model can be established. On the other hand, the disadvantages of the KNN algorithm are as follows. Researchers should arbitrarily set the number of optimal neighbors (k) and the distance scale to be used to suit the characteristics of each data. And it depends on the data, so it should be found through Grid Search. At this time, the distance between the new observations and each learning data must be measured in full, so the computation takes a long time. To mitigate this, Locality Sensitive Hashing, Network based Indexer, and Optimized Product Quantization are typically attempted.

First, you have assigned label values for three regions. I then randomly selected the extracted ENF data using a random function, and extracted the coefficients for the prediction using the methods described in the previous experimental section. You have specified the y-values and train data that are based on this. Next, I set the training data as input values for the support vector machine and KNN, respectively. Furthermore, as I proceed with each model, the prediction algorithm of each technique compares the characteristics of previously learned values to generate appropriate local label values for each input value. And now, in the next step, I need to use this process to understand how different probabilistic the predicted label values are extracted from the actual values.

7 Results

As a result of applying ENF signals from each region to ARIMA models, coefficient values were extracted for each region. I also visually checked whether these coefficients exhibit significant properties different from region to region, but I used machine learning classification techniques to determine whether they actually have significant properties. As a result, the coefficients in each domain had properties that enabled accurate classification.

It is necessary to classify it correctly and ensure that it is correct to proceed with the prediction on new data that has not been subsequently learned. For this reason, I learn regional characteristics, then perform a prediction step on specific data and perform a procedure to validate them. At this time, the trains and tests required for verification are set at a ratio of 8:2 before learning.

Unlike regression, the classification problem requires a variety of performance evaluation criteria. Thus, in this paper, I use the 'F1 score' of the classification performance evaluation model. The simplest way to measure performance is to calculate the accuracy. Accuracy is the number of correctly predicted data divided by the total number of data. The formula is as follows. Here's what each term means in the grade prediction classifier:

Confusion matrix (count)		*Predicted class*	
		Positive	Negative
Actual class	Positive	T P	F N
	Negative	F P	T N

Table 1. Confusion Matrix

- **True Positives**: the model predicted that the grade would be higher than B, and if so,
- **True Negatives**: the model predicted that the performance would be lower than B, and in practice, that would be the case.
- **False positives**: the model predicted that the grade would be higher than B and not actually
- **False Negatives**: the model predicted that the performance would be lower than B and not in practice

However, the classification model should not only be accurate, but also consider Recall. This is because accuracy can also represent very incorrect statistics depending on the data. Suppose you always create a model that predicts false results regardless of what the data is, this model has surprisingly high accuracy. However, despite having a high accuracy, this model is useless. In this situation, the statistics that can help are recall. Recall is the number of data that the model recognized as True that the data was actually True. And furthermore, I need to consider Precision. Because recall is also not a perfect statistic. The accuracy may be low, but if the model predicts all days as true, recall is 1. This model is no different from the one always predicted by False, but it has a high recall. In this situation, the statistics that can help are precision. Precision is the number of data that the model predicts to be true that is actually true.

That is, if the data label is an imbalanced structure, I decided to evaluate the classification by F1 score because it requires indicators that can reduce bias to accurately evaluate the performance of the model, and it is more important to consider TP+TN for the study purposes in this paper. These confusion matrices are shown in the Table 1.

If you refer to the table, the F1 score is calculated as shown in (3). Based on this, the F1 score for each predicted data predicted through SVM and K-neighbors classification techniques is shown in Table 2.

$$F1 = 2 \cdot \frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2 \cdot \frac{precision - recall}{precision + recall}. \tag{4}$$

Train Group	**F1 Score**	
	Support Vector Machine	*K-neighbors Classifier*
West	0.8	0.75
East	0.7	0.75
Hawaii	0.9	0.85

Table 2. Accuracy Score of Training and Predicting Data

As shown in Table 2, comparing the results of local labeling prediction for SVM techniques and K-neighbor classifier techniques, both showed similar accuracy for regions. Furthermore, local prediction techniques using ARIMA coefficients were able to clearly confirm that the area where the data originated could be accurately found without interpolation.

8 Conclusion

As information technology advances, an increasing number of digital data is being used in too many ways. Along with this, malicious use of voice and video data that can be easily used online is also increasing. This paper proposes a novel digital forensics technique that uses ARIMA to find the characteristics of ENF data collected from online multimedia and utilize them for prediction to track locations more efficiently and simply than interpolation. Digital forensics using grid frequency signal data varies, but in this paper, characteristics of grid frequency signals unique to each region are identified by coefficients. To obtain these coefficients, it is important to estimate the correct order for the data. To extract coefficients meaning local characteristics and learn to predict them, I conduct experiments in ARIMA in which data are grouped into specific sizes and learned iteratively instead of simply learning time series data as models at once. I then proceed to classify the extracted coefficient values using machine learning techniques. In addition to the data applied to classification learners like this, I also create a model that derives a given label value for the new data.

Research implications: academic implications, practical (technical/policy) implications is this: It is important that accurate location detection can be achieved efficiently without additional interpolation if the purpose is simply to find the grid where the signal originated through voice/video data. Furthermore, I confirmed that it has significant results accuracy compared to interpolation. It is expected that accurate location detection research will be conducted not only in the U.S. but also around the world by utilizing ENF data extracted from online multimedia videos in the future. I also think it is meaningful in terms of technology. Recently, artificial neural networks have been widely used in time series analysis along with trends such as deep learning and artificial intelligence. Specifically, RNN, LSTM, etc. are available for time series analysis, but the biggest drawback of neural network models is (so far)The point is that there is little proper model interpretation. Only algorithmically can perform well, but in data science, especially in a decision support system, the final decision maker is a human, so the model needs the basis for such prediction. From this perspective, classical statistical-based time series analysis methods such as AR, MA, etc. are judged to be more suitable for the topic and purpose of this paper.

9 Limitations and Future Work

9.1 Limitations of Research

The proposed technique has several limitations.

- Limitations of ENF signals themselves The proposed positioning technique in the paper essentially extracts ENF signals based on speech signals from images and proceeds with predictions based on them. However, ENF signals have variability that can be changed for fundamentally different reasons. In addition, ENF signals pose a problem that data may not be collected intact when collected because they originate from power grids. Generally, methods to extract ENF signals using FDRs also proceed with collection using median filters to eliminate the noon difference, as the nature of these ENF signals can reduce data throughput or cause high noise when used with measurement errors. The paper also has a lesser degree of problem with these problems than when collecting signals directly using FDR, but nevertheless has limitations that it will be affected to some extent by the variability of ENF signals.
- Fine-grained grid measurements The proposed positioning technique predicts data excluding time data, which can lead to limitations when drawn on maps. At the intragrid level, most existing studies have assumed that ENF signals across interconnected power grids are simultaneously similar. However, due to the local variations in the load on the grid and the finite propagation rate of the load variations on different parts of the grid, minor variations in the frequency variations at different locations are likely to exist. Moreover, there may be

concerns that these problems could result in less accurate predictions if the proposed positioning technique is used in the paper.

9.2 Future Work

Through experiments so far, this paper collects images within online multimedia services and extracts ENF signals from speech data within images. Furthermore, the unique ENF signal coefficient value for each region is confirmed by the ARIMA model. Based on these experiments, I propose a positioning technique through images.

In other words, I used the properties of such power grids to predict ENF signals by analyzing online services. Power grids have patterns similar to periodicity. To validate our method, I need to analyze more data from other power grids. So I have to identify the possibility of generalizing our proposed method.

References

Abe, Mototsugu and Julius O Smith III. 2004. "Design criteria for the quadratically interpolated FFT method (I): Bias due to interpolation." *CCRMA STAN-M-114* .

Box, GEORGE EP, Gwilym M Jenkins and G Reinsel. 1970. "Time series analysis: forecasting and control Holden-day San Francisco." *BoxTime Series Analysis: Forecasting and Control Holden Day1970* .

Cheung, Yin-Wong and Kon S Lai. 1995. "Lag order and critical values of the augmented Dickey–Fuller test." *Journal of Business & Economic Statistics* 13(3):277–280.

Foody, Giles M and Ajay Mathur. 2004. "Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification." *Remote Sensing of Environment* 93(1-2):107–117.

Garg, Ravi, Adi Hajj-Ahmad and Min Wu. 2013. Geo-location estimation from electrical network frequency signals. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE pp. 2862–2866.

Garg, Ravi, Avinash L Varna and Min Wu. 2011. " Seeing" ENF: natural time stamp for digital video via optical sensing and signal processing. In *Proceedings of the 19th ACM international conference on Multimedia*. pp. 23–32.

Grigoras, Catalin. 2007. "Applications of ENF criterion in forensic audio, video, computer and telecommunication analysis." *Forensic science international* 167(2-3):136–145.

Kim, Hyunsoo, Youngbae Jeon and Ji Won Yoon. 2017. Construction of a national scale ENF map using online multimedia data. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. pp. 19–28.

Kohn, Robert and Craig F Ansley. 1986. "Estimation, prediction, and interpolation for ARIMA models with missing data." *Journal of the American statistical Association* 81(395):751–761.

Schmidt, Ralph. 1986. "Multiple emitter location and signal parameter estimation." *IEEE transactions on antennas and propagation* 34(3):276–280.