Secure Reversible Image Data Hiding over Encrypted Domain via Key Modulation

Dr. J. Kolangiappan.
M.Sc., M.Phil., B.Ed., Ph.D., Head & Assistant professor, PG & Research Department of Computer Science

Abstract: This work proposes a novel reversible image data hiding (RIDH) scheme over encrypted domain. The data embedding is achieved through a public key modulation mechanism, in which access to the secret encryption key is not needed. At the decoder side, a powerful two-class SVM classifier is designed to distinguish encrypted and non-encrypted image patches, allowing us to jointly decode the embedded message and the original image signal. Compared with the state-of-the-arts, the proposed approach provides higher embedding capacity, and is able to perfectly reconstruct the original image as well as the embedded message. Extensive experimental results are provided to validate the superior performance of our scheme.

Keywords: *HS-Histogram Shifting; PEE - Prediction Error Expansion*

I. INTRODUCTION

Reversible image data hiding (RIDH) is a special category of data hiding technique, which ensures perfect reconstruction of the cover image upon the extraction of the embedded message. The reversibility makes such image data hiding approach particularly attractive in the critical scenarios, e.g., military and remote sensing, medical images sharing, law forensics and copyright authentication, where high fidelity of the reconstructed cover image is required. The majority of the existing RIDH algorithms are designed over the plaintext domain, namely, the message bits are embedded into the original, un-encrypted images.

The early works mainly utilized the lossless compression algorithm to compress certain image features, in order to vacate room for message embedding. However, the embedding capacity of this type of method is rather limited and the incurred distortion on the watermarked image is severe. Histogram shifting (HS)- based technique, initially designed by Ni *et al.* Another class of approach achieving better embedding performance through shifting the histogram of some image features. The latest difference expansion (DE)-based schemes and the improved prediction error expansion (PEE)-based strategies were shown to be able to offer the state-of-the-art capacity distortion performance.

Recently, the research on signal processing over encrypted domain has gained increasing attention, primarily driven by the needs from Cloud computing platforms and various privacy preserving applications. This has triggered the investigation of embedding additional data in the encrypted images in a reversible fashion. In many practical scenarios, e.g., secure remote sensing and Cloud computing, the parties who process the image data are un-trusted.

To protect the privacy and security, all images will be encrypted before being forwarded to a un-trusted third party for further processing. For instance, in secure remote sensing, the satellite images, upon being captured by on-board cameras, are encrypted and then sent to the base station(s). After receiving the encrypted images, the base station embeds a confidential message, e.g., base station ID, location information, time of arrival (TOA), local temperature, wind speed, etc., into the encrypted images.

For security reasons, any base station has no privilege of accessing the secret encryption key K pre-negotiated between the satellite and the data center. This implies that the message embedding operations have to be conducted entirely over the encrypted domain. In addition, similar to the case of Cloud computing, it is practically very costly to implement a reliable key management system (KMS) in such multi-party environment over insecure public networks, due to the differences in ownership and control of underlying infrastructures on which the KMS and the protected resources are located.

It is therefore much desired if secure data hiding could be achieved *without* an additional secret data hiding key shared between the base station and the data center. Also, we appreciate simple embedding algorithm as the base station usually is constrained by limited computing capabilities and/or power. Finally, the data center, which has abundant computing resources, extracts the embedded message and recovers the original image by using the encryption key K. In this work, we propose an encrypted-domain RIDH scheme by specifically taking the above-mentioned design preferences into consideration.

The proposed technique embeds message through a public key modulation mechanism, and performs data extraction by exploiting the statistical distinguish ability of encrypted and non-encrypted image blocks. Since the decoding of the message bits and the original image is tied together, our proposed technique belongs to the category of *non-separable* RIDH solutions. Compared with the state-of-the-arts, the proposed approach provides higher embedding capacity, and is able to achieve perfect reconstruction of the original image as well as the embedded message bits. Extensive experimental results on 100 test images validate the superior performance of our scheme. The rest of this paper is organized as follows. Briefly overviews the related work on RIDH over the encrypted domain. Presents the proposed data hiding technique in encrypted images, we describe the approach for data extraction by exploiting the statistical distinguishes ability of encrypted and non-encrypted image blocks.

II. LITERATURE SURVEY

IMAGE PROCESSING

In imaging science, **Image Processing** is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Images are also processed as threedimensional signals with the third-dimension being time or the z-axis.

Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The *acquisition* of images (producing the input image in the first place) is referred to as imaging.

Closely related to image processing are computer graphics and computer vision. In computer graphics, images are manually *made* from physical models of objects, environments, and lighting, instead of being acquired (via imaging devices such as cameras) from *natural* scenes, as in most animated movies. Computer vision, on the other hand, is often considered *high-level* image processing out of which a machine/computer/software intends to decipher the physical contents of an image or a sequence of images (e.g., videos or 3D full-body magnetic resonance scans).

In modern sciences and technologies, images also gain much broader scopes due to the ever growing importance of scientific visualization (of often large-scale complex scientific/experimental data). Examples include microarray data in genetic research, or real-time multi-asset portfolio trading in finance.

DIGITAL IMAGE PROCESSING

This article is about mathematical processing of digital images. For artistic processing of images, see Image editing.

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of multidimensional systems.

HISTORY

Many of the techniques of digital image processing, or digital picture processing as it often was called, were developed in the 1960s at the Jet Propulsion Laboratory, Massachusetts Institute of Technology, Bell Laboratories, University of Maryland, and a few other research facilities, with application to satellite imagery, wire-photo standards conversion, medical imaging, videophone, character recognition, and photograph enhancement.[1] The cost of processing was fairly high, however, with the computing equipment of that era.

That changed in the 1970s, when digital image processing proliferated as cheaper computers and dedicated hardware became available. Images then could be processed in real time, for some dedicated problems such as television standards conversion. As general-purpose computers became faster, they started to take over the role of dedicated hardware for all but the most specialized and computer-intensive operations.

With the fast computers and signal processors available in the 2000s, digital image processing has become the most common form of image processing and generally, is used because it is not only the most versatile method, but also the cheapest.

Digital image processing technology for medical applications was inducted into the Space Foundation Space Technology Hall of Fame in 1994.

In 2002 Raanan Fattel, introduced Gradient domain image processing, a new way to process images in which the differences between pixels are manipulated rather than the pixel values themselves.

STATISTICAL CLASSIFICATION

In machine learning and statistics, **classification** is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. An example would be assigning a given email into "spam" or "non-spam" classes or assigning a diagnosis to a given patient as described by observed characteristics of the patient (gender, blood pressure, presence or absence of certain symptoms, etc.). Classification is an example of pattern recognition.

In the terminology of machine learning, classification is considered an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available. The corresponding unsupervised procedure is known as clustering, and involves grouping data into categories based on some measure of inherent similarity or distance.

Often, the individual observations are analyzed into a set of quantifiable properties, known variously as explanatory variables or *features*. These properties may variously be categorical (e.g. "A", "B", "AB" or "O", for blood type), ordinal (e.g. "large", "medium" or "small"), integer-valued (e.g. the number of occurrences of a particular word in an email) or real-valued (e.g. a measurement of blood pressure) Other classifiers work by comparing observations to previous observations by means of a similarity or distance function.

An algorithm that implements classification, especially in a concrete implementation, is known as a **classifier**. The term "classifier" sometimes also refers to the mathematical function, implemented by a classification algorithm, that maps input data to a category.

Terminology across fields is quite varied. In statistics, where classification is often done with logistic regression or a similar procedure, the properties of observations are termed explanatory variables (or independent variables, regressors, etc.), and the categories to be predicted are known as outcomes, which are considered to be possible values of the dependent variable. In machine learning, the observations are often known as *instances*, the explanatory variables are termed *features* (grouped into a feature vector), and the possible categories to be predicted are *classes*. Other fields may use different terminology: e.g. in community ecology, the term "classification" normally refers to cluster analysis, i.e. a type of unsupervised learning, rather than the supervised learning described in this article.

FEATURE EXTRACTION

In machine learning, pattern recognition and in image processing, **feature extraction** starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations. Feature extraction is related to dimensionality reduction.

When the input data to an algorithm is too large to be processed and it is suspected to be redundant (e.g. the same measurement in both feet and meters, or the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named a feature vector). Determining a subset of the initial features is called *feature selection*. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.

Feature extraction involves reducing the amount of resources required to describe a large set of data. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power, also it may cause a classification algorithm to overfit to training samples and generalize poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

The best results are achieved when an expert constructs a set of application-dependent features, a process called feature engineering. Nevertheless, if no such expert knowledge is available, general dimensionality reduction techniques may help.

These include:

- Independent component analysis
- Isomap
- Kernel PCA
- Latent semantic analysis
- Partial least squares
- Principal component analysis
- Multifactor dimensionality reduction
- Nonlinear dimensionality reduction
- Multilinear Principal Component Analysis
- Multilinear subspace learning
- Semidefinite embedding
- Autoencoder
- Deep feature synthesis

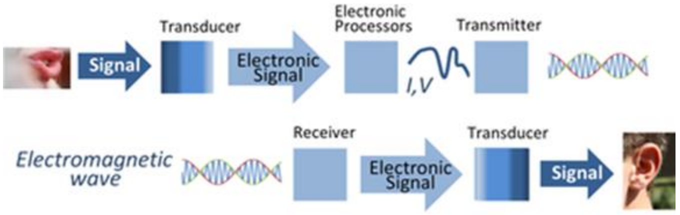
IMAGE PROCESSING

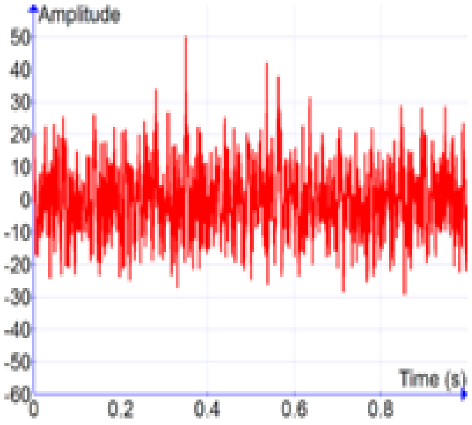
One very important area of application is image processing, in which algorithms are used to detect and isolate various desired portions or shapes (features) of a digitized image or video stream. It is particularly important in the area of optical character recognition.

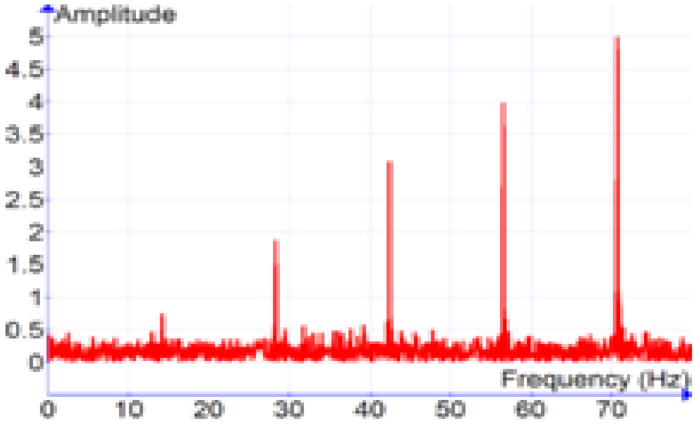
SIGNAL PROCESSING

"Signal theory" redirects here. It is not to be confused with Signalling theory or Signalling (economics).

Signal processing is an enabling technology that encompasses the fundamental theory, applications, algorithms, and implementations of processing or transferring information contained in many different physical, symbolic, or abstract formats broadly designated as *signals*. It uses mathematical, statistical, computational, heuristic, and linguistic representations, formalisms, and techniques for representation, modelling, analysis, synthesis, discovery, recovery, sensing, acquisition, extraction, learning, security, or forensics







APPLICATION FIELDS

- Audio signal processing – for electrical signals representing sound, such as speech or music
- Digital signal processing
- Speech signal processing – for processing and interpreting spoken words
- Image processing – in digital cameras, computers and various imaging systems
- Video processing – for interpreting moving pictures
- Wireless communication – waveform generations, demodulation, filtering, equalization
- Control systems
- Array processing – for processing signals from arrays of sensors
- Process control – a variety of signals are used, including the industry standard 4-20 mA current loop
- Seismology
- Financial signal processing – analyzing financial data using signal processing techniques, especially for prediction purposes.
- Feature extraction, such as image understanding and speech recognition.
- Quality improvement, such as noise reduction, image enhancement, and echo cancellation.
- (Source coding), including audio compression, image compression, and video compression.
- Genomics, Genomic signal processing [3]
- In communication systems, signal processing may occur at:

- OSI layer 1 in the seven layer OSI model, the Physical
  Layer (modulation, equalization, multiplexing, etc.);
- OSI layer 2, the Data Link Layer (Forward Error Correction);
- OSI layer 6, the Presentation Layer (source coding, including analog-to-digital conversion and signal compression).

PATTERN RECOGNITION

Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and regularities in data, although it is in some cases considered to be nearly synonymous with machine learning. Pattern recognition systems are in many cases trained from labeled "training" data (supervised learning), but when no labeled data are available other algorithms can be used to discover previously unknown patterns (unsupervised learning).

The terms pattern recognition, machine learning, data mining and knowledge discovery in databases (KDD) are hard to separate, as they largely overlap in their scope. Machine learning is the common term for supervised learning methods and originates from artificial intelligence, whereas KDD and data mining have a larger focus on unsupervised methods and stronger connection to business use.

Pattern recognition has its origins in engineering, and the term is popular in the context of computer vision: a leading computer vision conference is named Conference on Computer Vision and Pattern Recognition. In pattern recognition, there may be a higher interest to formalize, explain and visualize the pattern, while machine learning traditionally focuses on maximizing the recognition rates.

Yet, all of these domains have evolved substantially from their roots in artificial intelligence, engineering and statistics, and they've become increasingly similar by integrating developments and ideas from each other.

In machine learning, pattern recognition is the assignment of a label to a given input value. In statistics, discriminate analysis was introduced for this same purpose in 1936. An example of pattern recognition is classification, which attempts to assign each input value to one of a given set of *classes* (for example, determine whether a given email is "spam" or "non-spam"). However, pattern recognition is a more general problem that encompasses other types of output as well.

Other examples are regression, which assigns a real-valued output to each input; sequence labeling, which assigns a class to each member of a sequence of values (for example, part of speech tagging, which assigns a part of speech to each word in an input sentence); and parsing, which assigns a parse tree to an input sentence, describing the syntactic structure of the sentence. Pattern recognition algorithms generally aim to provide a reasonable answer for all possible inputs and to perform "most likely" matching of the inputs, taking into account their statistical variation.

This is opposed to *pattern matching* algorithms, which look for exact matches in the input with pre-existing patterns. In contrast to pattern recognition, pattern matching is generally not considered a type of machine learning, although pattern-matching algorithms (especially with fairly general, carefully tailored patterns) can sometimes succeed in providing similar-quality output of the sort provided by pattern-recognition algorithms.

CONCLUSION

In this paper, we design a secure reversible image data hiding (RIDH) scheme operated over the encrypted domain. We suggest a public key modulation mechanism, which allows us to embed the data via simple XOR operations, without the need of accessing the secret encryption key. At the decoder side, we propose to use a powerful two-class SVM classifier to discriminate encrypted and non-encrypted image patches, enabling us to jointly decode the embedded message and the original image signal perfectly. We also have performed extensive experiments to validate the superior embedding performance of our proposed RIDH method over encrypted domain.

References

[1] Ahmed, S. and H.R. Warren, 1989. The Radarsat System. *IGARSS'89/12th Canadian Symposium on Remote Sensing.* Vol. 1. pp.213-217.

[2] Anger, C.D., S. K. Babey, and R. J. Adamson, 1990, A New Approach to Imaging Spectroscopy, SPIE Proceedings, *Imaging Spectroscopy of the Terrestrial Environment,* 1298: 72 - 86. - specifically, CASI

[3] Curlander, J.C., and McDonough R. N., 1991. *Synthetic Aperture Radar, Systems & Signal Processing*. John Wiley and Sons: New York.

[4] Elachi, C., 1987. *Introduction to the Physics and Techniques of Remote Sensing*. John Wiley and Sons, New York.

[5] King, D., 1992. Development and application of an airborne multispectral digital frame camera sensor. XVIIth Congress of ISPRS, *International Archives of Photogrammetry and Remote Sensing.* B1:190-192.

[6] Lenz, R. and D. Fritsch, 1990. Accuracy of videometry with CCD sensors. *ISPRS Journal of Photogrammetry and Remote Sensing,* 90-110.

[7] Lillesand, T.M. and Kiefer, R.W., 1994, *Remote Sensing and Image Interpretation*, 3rd. Ed., John Wiley and Sons, Inc.: Toronto.

[8] Luscombe, A.P., 1989. The Radarsat Synthetic Aperture Radar System. *IGARSS'89/12th Canadian Symposium on Remote Sensing.* Vol. 1. pp.218-221.

[9] Staenz, K., 1992. A decade of imaging spectrometry in Canada. *Canadian Journal of Remote Sensing*. 18(4):187-197.

[10] Jensen, J.R., 1986. *Digital Image Processing, a Remote Sensing Perspective*.

[11] Schwarz, K-P., Chapman, M.A., Canon, E.C. and Gong, P., 1993. An integrated INS/GPS approach to the georeferencing of remotely sensed data. *Photogrametric Engineering and Remote Sensing,* 59(11): 1667-1673.

[12] Shlien, S., 1979. Geometric correction, registration, and resampling of Landsat Imagery. *Canadian Journal of Remote Sensing*. 5(1):74-87.

[13] Forster, B.C., 1984. Derivation of atmspheric correction procedures for Landsat MSS with particular reference to urban data. *Int. J. of Remote Sensing* . 5(5):799-817.

[14] Horn, B.K.P., 1986. *Robot Vision.* The MIT Press:Toronto.

[15] Horn, B.K.P., and Woodham, R.J., 1979. Destriping Landsat MSS images by histogram modification. *Computer Graphics and Image Processing*. 10:69-83.

[16] Richards, J.A., 1986. *Digital Image Processing*. Springer-Verlag: Berlin.

[17] Tanre, D., Deuze, J.L., Herman, M., Santer, R., Vermonte, E., 1990. Second simulation of the satellite signal in the solar spectrum - 6S code. *IGARSS'90*, Washington D.C., p. 187.