Journal of Multimedia

ISSN 1796-2048

Volume 7, Number 4, August 2012

Contents

Special Issue: Multimedia Contents Security in Social Networks Applications
Guest Editors: Zhiyong Zhang and Muthucumaru Maheswaran

Guest Editorial
Zhiyong Zhang and Muthucumaru Maheswaran

SPECIAL ISSUE PAPERS

DRTEMBB: Dynamic Recommendation Trust Evaluation Model Based on Bidding
Gang Wang and Xiao-lin Gui

Block-Based Parallel Intra Prediction Scheme for HEVC
Jie Jiang, Baolong, Wei Mo, and Kefeng Fan

Optimized LSB Matching Steganography Based on Fisher Information
Yi-feng Sun, Dan-mei Niu, Guang-ming Tang, and Zhan-zhan Gao

A Novel Robust Zero-Watermarking Scheme Based on Discrete Wavelet Transform
Yu Yang, Min Lei, Huaqun Liu, Yajian Zhou, and Qun Luo

Stego Key Estimation in LSB Steganography
Jing Liu and Guangming Tang

REGULAR PAPERS

Facial Expression Spacial Charts for Describing Dynamic Diversity of Facial Expressions
H. Madokoro and K. Sato
Guest Editorial

Recent years have witnessed rapid developments of the information and communication technologies, and the Next-Generation Internet, 3G and 4G wireless mobile networks have been striding to a large-scale deployment and application. The emerging versatile network admissions, especially for social network services, tools and applications, enable a much more convenient access to multimedia contents in anytime, at anywhere, for anyone. Unfortunately, in such environment, copyright infringement behaviors, such as illicit copying, malicious distribution, unauthorized usage, free sharing of copyright-protected digital contents, will also become a much more common phenomenon. Some research frontiers on multimedia contents security in social networks applications have been in progress, including enhanced security mechanisms, methods and algorithms, trust assessment and risk management in social network applications, as well as social factors and soft computing in social media distributions.

The special issue attempts to bring together researchers, contents industry engineers and administrators resorting to the state-of-the-art technologies and ideas to protect valuable multimedia contents and services against attacks and IP piracy in the emerging social networks. It includes 5 selected papers as follows:

The first paper focuses on an interesting recommendation trust issue in the large-scale distributed computing, including social networks. Dr. Gang Wang proposed a dynamic recommendation trust evaluated model based on bidding in E-Commerce environment. By greatly increasing the “criminal cost” of malicious recommendation nodes, with recommendation optimization algorithm based on Markov chain, the model ensures that the recommendation service is true and more objective, which will stimulate the enthusiasm of nodes objective recommendation and reach the goal that nodes give up malicious recommendation and cooperative cheating on its own initiative by objective method so as to restrain malicious recommendation and cooperative cheating effectively. The model also shows good restraint on malicious recommendation by a serious of simulation experiments.

Then, the continuous emergence of multimedia video coding standards is cared in the second paper. Especially, the existing challenges are two folds: one fold is to find efficient coding algorithms which require high performance, and the other is to speed up the coding process. With recent advancement of VLSI (the Very Large Scale Integration) semiconductor technology contributing to the emerging digital multimedia word, this paper intends to investigate efficient parallel architecture for the emerging high efficiency video coding (HEVC) standard to speed up the intra coding process, without any prediction modes ignored. Dr. Jie Jiang’s experimental implementations of the proposed algorithm are demonstrated by using a set of video test sequences that are widely used and freely available. The results show that the proposed algorithm can achieve a satisfying intra parallelism without any significant performance loss.

In the third paper, LSB matching steganography technologies were explored and discussed in detail. The authors proposed a novel optimized LSB matching steganography scheme based on Fisher Information. The embedding algorithm is designed to solve the optimization problem, in which Fisher information is the objective function and embedding transferring probabilities are variables to be optimized. By modeling the groups of elements in a cover image as Gaussian mixture model, the joint probability distribution of cover elements for each cover image is obtained by estimating the parameters of Gaussian mixture distribution. Finally, in order to embed message bits, pixels chose to add or subtract one according to the optimized transferring probabilities of the category. The experiments show that the security performance of this new algorithm is better than the existing LSB matching.

As one of key technologies for digital rights management, the research on watermarking algorithms is followed. In this paper, authors denoted the unsolved issues on traditional watermarking algorithms. For instance, the insertion of watermark into the original signal inevitably introduces some perceptible quality degradation. Another problem is the inherent conflict between imperceptibility and robustness. Some existing zero-watermarking algorithm available for audio and image cannot resist against some signal processing manipulations or malicious attacks. In the paper, a novel audio zero-watermarking scheme based on discrete wavelet transform (DWT) is proposed, which is more efficient and robust. The experiments show that the algorithm is robust against the common audio signal processing operations such as MP3 compression, re-sampling, low-pass filtering, cutting-replacement, additive white Gaussian noise and so on. These results demonstrate that the proposed watermarking method can be a suitable candidate for audio copyright protection.

In the last paper of the special issue, a burning issue on the hiding harmful information in multimedia conveniently in the emerging multimedia social networks. Dr. Jing Liu considered the problem of estimating the stego key used for hiding using least significant bit (LSB) paradigm, which has been proved much difficult than detecting the hidden message. Previous framework for stego key search was provided by the theory of hypothesis testing. However, the test threshold is hard to determine. In this paper, we propose a new method, in which the correct key can be identified by inspecting the difference between stego sequence and its shifted sequence on the embedding path. It’s shown that the new technique is much simpler and quicker than that previously known.

The above five paper were selected by strict two rounds of peer reviewing based on their originality, relevance, technical clarity and presentation, by at least two anonymous reviewers. Here, I show gratitude to Prof. Maheswaran for...
his collaboration on the successful special issue on the very interesting and challenging topic of multimedia contents security in social networks. Besides, all invited reviewers are appreciated for their reviewing, comments and suggestions for authors. Finally, I give special thanks to Prof. Jiebo Luo and Dr. George Sun for their great helps and efforts so as to the special issue publication on time and successfully.

Guest Editors

Zhiyong Zhang (Email: z.zhang@ieee.org)
Department of Computer Science, Henan University of Science & Technology, Luoyang, P. R. of China

Muthucumaru Maheswaran (Email: maheswar@cs.mcgill.ca)
School of Computer Science, McGill University, Montreal, Canada

Zhiyong Zhang, received his Master, Ph.D. degrees in Computer Science from Dalian University of Technology and Xidian University, China, respectively, and post-doctoral fellowship at Xi'an Jiaotong University, China.

He is currently associate professor with College of Electronics Information Engineering, Henan University of Science & Technology, and research interests include digital rights management and soft computing, trusted computing and access control, as well as security risk management. Recent years, he has published over 50 scientific papers and 2 monographs on the above research fields, and held 3 authorized patents.


Muthucumaru Maheswaran, received BSc degree in Electrical and Electronic Engineering from University of Peradeniya, Sri Lanka in 1990. In 1994, he received a MSEE degree in Electrical Engineering from School of Electrical and Computer Engineering at Purdue University, West Lafayette, Indiana, USA. From the same school he also received a PhD in Computer Engineering in December 1998.

Prof. Maheswaran is nowadays an associate professor at the School of Computer Science and Department of Electrical and Computer Engineering of McGill University, Canada. He is directing the research activities at the Advanced Networking Research Lab, and His research interests are in the general areas of Computer Networking, Information Security, and Distributed Systems. He has published hundreds of scientific on the related research realms in numerous international journals and conferences.
DRTEMBB: Dynamic Recommendation Trust Evaluation Model Based on Bidding

Gang Wang\textsuperscript{1,2,3}
\textsuperscript{1}Department of Electronics and Information Engineering, Xi’an Jiaotong University, Xi’an, China
\textsuperscript{2}Shanxi Province Key Laboratory of Computer Network, Xi’an Jiaotong University, Xi’an, China
\textsuperscript{3}School of Information, Xi’an University of Finance and Economics, Xi’an, China

Wgang_only@126.com

Xiao-lin Gui\textsuperscript{1,2}
\textsuperscript{1}Department of Electronics and Information Engineering, Xi’an Jiaotong University, Xi’an, China
\textsuperscript{2}Shanxi Province Key Laboratory of Computer Network, Xi’an Jiaotong University, Xi’an, China

xlgui@mail.xjtu.edu.cn

Abstract—Fraud or cheating of malicious nodes was restrained by reducing the trust value of malicious nodes in the past trust model, but few recommendation nodes for cooperation cheating were punished, so it has had only limited malicious recommendation or fraudulent action against again and again. Due to lack of punishment for malicious recommendation of recommendation nodes, enormous malicious recommendation nodes remain and search for next criminal opportunity. By borrowing idea of bidding and the introduction of competition and the mechanism of rewards and penalty, this paper proposes a dynamic recommendation trust evaluated model based on bidding in E-Commerce environment. By greatly increasing the “criminal cost” of malicious recommendation nodes, with recommendation optimization algorithm based on Markov chain, the model ensures that the recommendation service is true and more objective, which will stimulate the enthusiasm of nodes objective recommendation and reach the goal that nodes give up malicious recommendation and cooperative cheating on its own initiative by objective method so as to restrain malicious recommendation and cooperative cheating effectively. In simulation experiment, the model also shows good restraint on malicious recommendation.

Index Terms—Bidding, Recommendation Trust, Evaluation Model, Recommendation Ability, Recommendation Oligarch

I. INTRODUCTION

With the rapid growth in computing and communications technology, the past decade has witnessed a proliferation of powerful e-commerce. However, the problems also follow it, in which, trust is a very important problem in mobile e-commerce. Thus, how to build a secure, reliable and trustworthy trust evaluation system for mobile e-commerce is a research hot spot nowadays.

In the large-scale distributed computing, because the both of services don’t know each other and lack of the necessary trust basis, there may be malicious node fraud, false trading and other problems. A better way is to interact with the high trust nodes by using good trust security model in a number of interactive objects. M. Blaze et al. proposed trust management concept in Literature [1] to solve the issue of trust in distributed environment at the first time, and depended on this to develop the corresponding trust management system, PolicyMaker and KeyNote in Literature [2]. Literature [3] proposes the idea of using bid to improve the efficiency of trust query of recommended nodes and the confidence updating mechanism, and uses mainly the subjective logic trust model that proposed by A. Jø Sang in Literature [4], which decides trust degree of object-evaluated through the introduction of uncertain factor for evaluation of experiment and by using three factors of credible, incredible and uncertainty. However, for the lack of effective incentive, the model can not promote effectively positive tender of the node. Literature [5] proposes an electronic community PeerTrust Model based on local reputation under P2P environment. This PeerTrust Model gives a comprehensive trust assessment based on the feedbacks acquired from nodes, the magnitude of the feedback, trust degree of the feedback, the critical and non-critical factors that differentiate the interactive context and the environmental factors related to the electronic community. Kamvar et al. in Stanford University in Literature [6] proposes the EigenRep model, which is a typical global reputation model, specifying the trust calculation under the P2P environment. Literature [7] proposes a global trust model that is intended to overcome the lack of security consideration of EigenRep Model, for instance, feigning, slandering and lack of punitive measures and solves the problem of trust.
recommendation effectively. Tang Wen and others proposed a subjective trust management model that is based on fuzzy set theory in literature [8], and the model builds a trust model by using fuzzy mathematics method for the ambiguity of subjective trust. Literature [9] classifies the trust values of nodes into resource trust value, node contribution value and node assessment value, but it ignores the fact that the trust value of a node is related to its performance with the influence of various factors. Literature [10] proposed a new trust model under the P2P E-Commerce environment, and the model accelerates direct trust by vote in order to stimulate nodes vote positively to increase the reliability of trust. Tian Chunqi and others proposed a novel Super-Peer based trust model for Peer-to-Peer Networks in literature [11]. Though it solves some trust problems, the quality of recommendation service (abbreviated as QoRS) could not be judged effectively, because they didn’t consider the uncertainty of recommendation information. Literature [12,13] summarized comprehensively the P2P trust mechanisms under the distributed computing environment, but because of the lack of research and analysis for collaborative cheating problems, it can not effectively reduce the emerging malicious recommendations issues in the network. Literature [14] proposed a trust model based on transaction content similarity, and the model represents the trust degree of recommendation by taking advantage of service contents similarity, and improves the reliability of trust recommendation. Due to the absence of appropriate incentive and punishment mechanism, there are still some limitations in reducing the malicious recommendation. Literature [15] proposed a new trust model based on advanced D-S evidence theory for P2P networks. Though the model improves the creditability of evaluation to node, the model could not effectively mobilize the node positivity and thus it is difficult to ensure healthy development of network because of lack of incentive and punishment mechanism for node. Literature [16] proposed dynamic trust evaluation model under distributed computing environment, which is based on the Dempster-Shafer Theory and Shapley entropy, and refers to the people’s trust relationship model of Sociology. The trust evaluation model uses historical interactive information to compute the direct trust. Shapley entropy evaluates the quantity of information of each node’s direct trust function, and then the direct trust is revised in consideration of each node’s reliability and its quantity of information. Afterwards, they compute the integrative trust derived from the revised direct trust of all nodes according to the Dempster rules. In addition to above, some researchers analyze the trust security model in terms of social network, such as Sabater and others, who proposed a reputation-based trust system REGRET [17], which uses social network analysis and a hierarchical ontology structure to integrate reputation of different types so as to calculate the final node trust value.

Through above analysis, although the past history trust models have reduced a few malicious services and fraud actions, these models are not good to restrain malicious recommendation actions of network nodes, we find the current trust models have the several questions:

1. Evaluation of the current trust models is to aim at both parties in transaction, but evaluation for recommendation nodes is little. So it is the difficult that trust models find malicious recommendation nodes.

2. Because QoRS is not only positive correlation with recommendation success ratio and reliability of recommendation nodes, but also positive correlation with improvement degree of recommendation ability. But the current evaluation method for QoRS is only to depend on transaction success or not, so evaluation method shows monotonous and inaccuracy, and evaluation effect is bad.

3. Because the current models’ incentive and penalty mechanism depends only on improvement or reduction of trust value for transaction nodes, these models lack to encourage the power for active recommendation of recommendation nodes.

To solve these issues, we proposes a dynamic recommendation trust evaluation model based on Bidding for E-Commerce environment in paper, and its novel contributions are as follows.

1. We propose a dynamic recommendation trust evaluation model based on bidding mechanism of e-commerce. Through bidding of recommendation nodes, many recommendation nodes can gain corresponding rewards from transactions, so our model can encourage a lot of recommendation nodes to recommend actively.

2. We propose a novel evaluation method for QoRS based on Markov. The evaluation method for QoRS integrates recommendation success ratio and the improvement degree of recommendation service ability. The method can effectively overcome monotonous and inaccuracy of evaluation, and solves effectively a problem that recommendation of node of the higher trust is trustworthy than nodes’ recommendation of the lower trust, and the method effectively avoids to arise “recommendation oligarch” in e-commerce network.

3. We distinguish recommendation nodes between acquaintance recommendation nodes and strange recommendation nodes, and distinguish acquaintance recommendation nodes between direct and indirect acquaintance recommendation nodes. So we can effectively ensure the trustworthy of recommendation according to acquaintance degree of recommendation nodes with evaluator.

This paper aims to build a dynamic trust evaluation model based on bidding idea for E-Commerce of under the social network environment. We summarize the related research work of the current trust model in section I. Section II introduces related key concepts and characteristics. Section III puts forward a novel dynamic recommendation trust evaluation model based on bidding. Section IV introduces evaluation method of recommendation service quality. Section V simulates the model by experiment and verified effective of proposing model. Section VI summarizes the paper and next research effort.
II. RELATED CONCEPT AND ATTRIBUTION

Definition 2.1. Trust is a type of faith and reliance, which reflects the subject’s degree of knowledge for the identity and behavioral intention of the object in a specific time and context.

Definition 2.2. Reputation, also known as prestige, is the collection of all recommendation value for an entity and a reflection of the trust that a subject has for the object.

Definition 2.3. Direct Trust, also known as local trust, refers to the judgment that the subject has for the object’s ability and trustworthiness and reliability, based on his/her experience that the subject acquires directly from the contact and transactions with the object.

Definition 2.4. Indirect Trust, also called recommendation trust, refers to the judgment of the subject on the object’s ability, trustworthiness and reliability through the recommendation of a third party.

The relationship of direct trust and recommendation trust is shown in Figure 1.

Figure 1. Local Trust and Recommendation Trust

Trust is a dualistic relationship, which can be one-to-one, one-to-many (or individual to group), many-to-one (group to individual) and many-to-many (or group to group) relationships. Trust is a complicated concept that bears both subjectivity and objectivity. This paper summarizes the following key attributes of trust:

1. Subjectivity

   Subjective trust is cognitive phenomenon of human beings about the real world and the subjective judgment of the subject on the special traits or behaviors of the object at a specific level. Different subjects have different subjective judgments at a specific object.

2. Conditional Transitivity

   Transitivity only takes place under specific conditions. For instance, entity A trusts entity B and entity B trusts entity C, but it cannot be inferred that entity A trusts entity C.

3. Asymmetry

   Entity A trusts Entity B, but entity B does not necessarily trusts entity A.

4. Context

   Trust is closely related to the context and the environment the subject is in. Under different contexts or backgrounds, the trust can be different.

5. Domain Correlation

   Trust can be represented differently pending changes of the domain and scope. For example, a medical doctor enjoys higher degree of trust in the medical field, but the degree of trust for him may be very low in the domain of music.

6. Content Correlation

   Entity A trusts entity B does not mean A trusts all behaviors of B. As a matter of fact, entity A trusts entity B only under a specific context for a particular type of behavior. For instance, A trusts B and is willing to lend the latter his/her bicycle, this is because that A believes that B will surely return the bicycle. On the other hand, A does not lend B ten thousand dollars because A has doubt whether B will be able to return ten thousand dollars.

Definition 2.5. Service Provider, also known target node, refers to the node which provides source service, and its trust will be evaluated by service requestor in e-commerce trust network, denoted by SP.

Definition 2.6. Service Recommendation node refers to the node which provides recommendation of service source for the service requestor in e-commerce trust network, in order that gains related economic interest and trust degree, denoted by SR.

Definition 2.7. Service requestor, also known Evaluator, refers to the node which evaluates trust of Service Provider in e-commerce trust network, denoted by E.

Definition 2.8. Recommendation Oligarch refers to a member of recommendation nodes whose recommendation will play the key role for recommendation of network, and will monopolize individual of right to recommend of the whole network.

III. DYNAMIC RECOMMENDATION TRUST EVALUATION MODEL.

To solve effectively the problem of collaborative cheating of malicious nodes, we need the effective assessment and management of the network nodes. From the point of view as social psychology and Organizational Behavioral Science, whether network nodes of participating service are good or malicious nodes, there are always with a certain purpose, which obtains corresponding benefits. Based on above analysis, by borrowing idea of bidding, the gains of normal service node is the higher than the Non-normal service one from long-term benefits. (This service includes two cases, one is to provide resource service, and the other is to provide a recommendation service). The benefits include two aspects: one is good nodes make a profit from their excellent recommendation and the other is the improvement of the good recommendation nodes credibility.

Definition 3.1. Trust evaluation management model, which contains three entities of service requestor (also known as evaluator), service provider(also knows as resource entities) and service recommendation nodes, is new network security model based on social psychology, which is built according to history interaction relation of three entities.

Definition 3.2. Direct acquaintance node is that those nodes that have already had direct interaction with the resource providers.

Definition 3.3. Indirect acquaintance node is that those nodes that have already had direct interaction with direct acquaintance node, and haven’t had direct interaction with evaluator.

Another recommendation node is stranger nodes except direct acquaintance nodes and indirect acquaintance nodes.

From social psychology, recommendation reliability and trustworthiness have relation to familiarity of between recommendation nodes and evaluator to a great extent. If there is experience of direct interaction between
recommendation nodes and evaluator, their recommendation reliability is the higher than recommendation of recommendation nodes of experience of non-direct interaction, and reliability of acquaintance recommendation is the higher than reliability of stranger recommendation, reliability of direct acquaintance recommendation is the higher than indirect acquaintance recommendation. In the e-commerce network environment, recommender is recommendation node, and evaluator is evaluation node, and objector evaluated is service providing node, also known target node. In which, recommendation node are classified into direct acquaintance node, indirect acquaintance node and strange recommendation node. Their relation is as following Figure2.

![Figure 2. Description for Relationship of Between Network Nodes](image)

In which, Service requestor is requestor of service, and service provider is the provider of service, and service recommendation nodes is a entity that recommends service for service requestor. While there are direct transaction experiences between service requestor and service provider, we call two parts have direct trust relation, on the contrary, there is indirect trust relation.

### A. The Basic Process for selecting Bidding Nodes

In dynamic recommendation trust model, the basic process of selecting bidding nodes is the first is evaluator to send bidding information; the second, service providing nodes select to bid according to bidding information, and bidding nodes are classified two categories: one is resource service providing nodes; the other is recommendation bidding nodes. After evaluator receives bidding application from service providing nodes, and evaluator queries trust value of bidding service nodes and evaluation bidding service nodes; the third is to select trust bidding nodes for contract and service.

**Definition 3.4. Service bidding messages is defined as following tetrad:**

\[
SBM = (Evaluator\_ID, ServiceContents, TimeStrap, Reward) \quad (1)
\]

In which, **Evaluator\_ID** is evaluator identity, that is service requestor identity, and **ServiceContents** is service contents of request for service requestor; **TimeStrap** is window of bidding time; **Reward** is that service provider and service recommendation nodes obtain the corresponding reward from evaluator, and **Reward** includes two sides: one is to obtain corresponding funds from evaluator; the other is to obtain trust evaluation from evaluator. Correspondingly, **Reward** represents to get corresponding punishment to false recommendation nodes and malicious service nodes.

**Definition 3.5. Resource service application is defined as following triple:**

\[
RSA = (Service\_Provider\_ID, ServiceContents, Response\_Time) \quad (2)
\]

In which, **Resource\_Provider\_ID** refers to identity of resource service bidding nodes; **ServiceContents** refers to service contents provided by **Service\_Provider\_ID**; **Response\_Time** refers to feedback time of bidding nodes.

**Definition 3.6. Recommendation service application is defined as following tetrad:**

\[
RSA' = (Recommendator\_ID, Service\_Provider\_ID, Recommendation\_Trust, Recommendation\_Response\_Time) \quad (3)
\]

In which, **Recommendator\_ID** refers to identity of recommendation service node; **Service\_Provider\_ID** refers to identity of resource provider; **Recommendation\_Trust** refers to recommendation trust, **Recommendation\_Response\_Time** refers to feedback time of recommendation nodes.

### B. Trust Computing

Trust computing consists of direct trust computing and recommendation trust computing. (Also known as indirect trust computing). Through integration of direct trust and recommendation trust, service requestor (or evaluator) can gain the global trust value of service provider. From research of literature[14], we find that there is more closely relationship in two factors to the global trust: one is the objectivity and accuracy of evaluation of coming from recommendation service nodes and service requestor to service provider, which depends on similarity of service contents between each evaluator and service provider and each recommendation service node and service provider in a period of time; the other is the global trust which depend on familiarity between service requestor and recommendation nodes. This relationship conform also social psychology, that is to say, the trustworthiness of acquaintance node evaluation is generally higher than that of stranger nodes, whereas the evaluation of direct acquaintance nodes is
comparatively more reliable than indirect acquaintance nodes.

(1) Direct trust computing

The so-called direct Trust $DT_{sp}^E$ refers direct trust value determined by the history of interactions between evaluator $E$ and service providing nodes.

$$DT_{sp}^E = \frac{S_{sp}^E + 1}{S_{sp}^E + F_{sp}^E + 2}$$  \hspace{1cm} (4)

$$G_{sp}^E = S_{sp}^E + F_{sp}^E$$  \hspace{1cm} (5)

In which, $DT_{sp}^E$ represent direct trust value, $S_{sp}^E$ represents the number of times that $E$ and $SP$ have been happy with the interaction, i.e., the success rate; $G_{sp}^E$ specifies the overall times of the transactions between $E$ and $SP$, $G_{sp}^E = S_{sp}^E + F_{sp}^E$, $F_{sp}^E$ refer the number of times that $E$ and $SP$ have been unhappy with the interaction, or rate of failures. when $G_{sp}^E = 0$, $DT_{sp}^E = 0.5$ $DT_{sp}^E$ is not equal with $DT_{sp}^E$.

Since trust is a dynamic Variable and has a characteristic of time decay, this paper introduces an influence function of time:

$$f(k) = \delta^{k - \frac{1}{n}}, 0 < \delta < 1$$  \hspace{1cm} (6)

There is the different influence in interaction information of the different time to trust computing, so there is a corresponding time impact factor in every time in order that node trust computing is the more accurate than history trust computing. In a given period of time $t_i$, success number $S_{sp}^E$ and failure number $F_{sp}^E$ with influence function of time $f(k)$ represent formula (6):

$$S_{sp}^E = S_{sp}^{E-1} + f(k) \times S_{sp}^{E-1} + F_{sp}^{E-1} + f(k) \times F_{sp}^{E-1}$$  \hspace{1cm} (7)

Finally, $S_{sp}^E$ and $F_{sp}^E$ are brought into formula (1), that is to compute direct trust $DT_{sp}^E$ of node $E$ to node $SP$ in $k$-th time

$$DT_{sp}^E = (S_{sp}^E + 1) / (S_{sp}^E + F_{sp}^E + 2)$$  \hspace{1cm} (8)

(2) general recommend trust computing

From Literature [14] obtains general recommend trust computing

$$RT_{sp} = \sum_{n=1}^{N} Sim_{ij} \cdot (\alpha \cdot \omega_{sp} \cdot DT_{sp} \cdot \beta \cdot \omega_{sp} \cdot DT_{sp}) / n$$  \hspace{1cm} (9)

$RT_{sp}$ is the total recommendation Trust Evaluation for node $SP$. $Sim_{ij}$ refers to the similarity of service content $C_{sp}$ and $C_{ij}$ between node $i$ to $SP$ and node $j$ to $SP$, $\omega_{sp}$ and $\omega_{sp}$ respectively refer to acquaintance recommendation weight and stranger recommendation weight; in which $\omega_{sp}$ and $\omega_{sp}$ : $\omega_{sp} = DT_{sp}^E, \omega_{sp} = 0.5$

$\alpha$ refers to the recommendation weight of direct acquaintance node or indirect acquaintance, and this represents the trust degree of the recommending node for the recommended node; $\alpha$ is set at:

$$\alpha = DT_{sp}^E, \beta = \sum_{k=1}^{N} PT_{sp} \cdot PT_{sp} \cdot PT_{sp} \cdot PT_{sp}$$  \hspace{1cm} (10)

Because there is a node cross situation in recommendation path of acquaintance nodes, let’s set a threshold $\epsilon$. When $\alpha$ = $\epsilon$, the recommendation path shall be abandoned. $\beta$ refers to stranger recommendation weight. If $\beta = \sum_{k=1}^{N} DT_{sp}^E$, while $\sum_{k=1}^{N} DT_{sp}^E = 0$, then the node is a newly joining node or a dormant node. In the process of trust recommendation, because there are two phenomena that recommendation path has the cross and independent, in which select a recommendation path of having highest trust value. In recommendation path shown by Figure 3, $A \rightarrow C \rightarrow E_i \rightarrow B$ will be selected.

![Figure 3. Recommendation Trust Path](image)

(3) service content similarity computing

This paper computes the similarity of service contents by vector included angle cosine method, and improves recommendation credibility; its computing formula is following:

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|}$$  \hspace{1cm} (11)

In which, $\vec{a}$ denotes service vector that service provider has provided for recommendation nodes in the period of time, and $\vec{b}$ denotes service vector that service provider will provide for evaluator this time, and their similarity is creditability of evaluator to recommendation node.

In which, $a_i = (x_{i1}, x_{i2}, \ldots, x_{in}), b_i = (x_{j1}, x_{j2}, \ldots, x_{jn})$ and they respectively refer to $Service$ $Contents$, $Response$ $Time$, $Service$ $Completion$, $Service$ $Cost$. Index set weight of Service vector has different situation at different distributed computing environment. After the above analysis, service similarity coefficient of this paper denotes as following:

$$Sim_{ij} = \frac{\sum_{i=1}^{n} \sigma_i a_{x_i} \cdot \sigma_i b_{x_j}}{\left(\sum_{i=1}^{n} \sigma_i a_{x_i}\right) \left(\sum_{i=1}^{n} \sigma_i b_{x_i}\right)^{1/2}}$$  \hspace{1cm} (12)

In which, $\sigma_i$ denote the weight of the $k$-th index.

(4) Global Trust

Global trust $GT_{sp}^E$ is to get trust value of service provider through fusion of recommendation trust and direct trust. Its computation equation is as follows:
In which, \( \lambda + \gamma = 1 \), \( \lambda, \gamma \in [0,1] \), \( \lambda \) is a weight factor of direct trust, and \( \gamma \) is a weight factor of recommendation trust. Because trust is a multi-attribute object, and is gradually attenuated with time, \( \lambda \) and \( \gamma \) is going to dynamic change with the numbers of interaction and others. While value of \( \lambda \) is larger and value of \( \gamma \) is smaller, proportion of direct trust will become larger and larger, and proportion of recommendation will become smaller and smaller as the numbers of interaction increase.

\[
\lambda = 1 - \left( \frac{1}{2} \right)^{\frac{1}{n-k}} \quad , \quad n-k \neq 0
\]  

(14)

C. Trust Update

The update of Trust value includes two aspects that one is the update of trust value of service provider, and the other is the update of trust value of the recommendation nodes. According to the final interaction situation and the global trust value, trust value of the service provider is updated positively (or negatively); and recommendation trust of each recommendation node is updated according to formula (14).

\[
\Delta = \left| RT_{E} - R_{oE} \right|
\]  

(15)

In which, \( RT_{E} \) represents general recommendation trust value of all recommendation nodes, and \( R_{oE} \) represents recommendation trust value of each recommendation node \( E \) to service provider \( SP \). \( \Delta \) is the difference between \( RT_{E} \) and \( R_{oE} \). We set update threshold \( \xi \), while \( \Delta \) is bigger than \( \xi \), it is the bigger that represents distance of general recommendation trust with recommendation trust of recommendation node \( i \), which recommendation node \( i \) is untrustworthy node; while \( \Delta \) is smaller than \( \xi \) or equivalent to \( \xi \), it is closer that represents distance of general recommendation trust with recommendation trust of recommendation node \( i \), and recommendation trust of recommendation node \( i \) is the more accurate, and recommendation node \( i \) is a trust node. \( \xi \) takes the average of all recommendation trust value.

\[
\xi = \frac{1}{n} \sum_{i=1}^{n} \xi_{i}
\]  

(16)

D. Model Algorithm Description

Input: Initialize network nodes, generate network nodes.

Output: a list of trusted nodes, a list of excellent recommendation nodes.

Step(1): service requestor (that is evaluator) publishes service bidding messages to network, then resource service providers receive messages to feed back its resource service bidding application.

Step(2): Resource service bidding nodes are fed back network and publicity, and recommendation service node will evaluate resource service bidding nodes and send to recommendation service bidding messages to network.

Step(3): According to demand conditions of recommendation service bidding, a few recommendation service bidding applications are cleaned by filter because they fall short of bidding request.

Step(4): Recommendation trust of each resource service bidding node is to be fused, in order to compute the general recommendation trust value;

Step(5): According to direct history interaction between resource service bidding node and evaluator, to compute the global trust value through fusion of direct trust and recommendation trust;

Step(6): While interactions are successful, there are a comparison between \( \Delta \) and threshold value \( \xi \) by using (4), while \( \Delta > \xi \), trust values is positive updated; else trust values is negative updated; and according to quality of recommendation service, a list of excellent recommendation nodes are produced automatically by this system, at the same time, and system notices evaluator pays fee to them.

IV. QoRS Evaluation Method

It is difficult to measure QoRS in the past historical trust model because the recommendation service ability of nodes is always developed and changed. Based on the analysis results, we find that the QoRS not only depends on the accuracy of the recommended trust service, but also depends on the degree of improvement of recommended service ability of the recommended node. The past history recommendation trust models often ignore improvement degree of recommendation service ability. We propose a dynamic recommendation evaluation method based on Markov chain, Abbreviated as DREM, the method not only considers accuracy of recommendation service, but considers improvement degree of recommendation service ability. The basic process for evaluation of QoRS is as follows:

Step1: To compute the recommendation accuracy of each trust recommendation node according to formula (15);

Step2: To compute improvement degree of recommendation ability of each recommendation node by DREMTBB;

Step3: To modify recommendation service quality value according to Step2, and get the list of excellent recommendation service nodes and the list of bad recommendation service nodes.

Markov chain is under the conditions that the state of process (or system) has known at time \( t_{0} \) and conditional distribution of the state of the process at time \( t > t_{0} \) is unrelated with the state of the process before at time \( t_{0} \), that is known to process under the conditions “now”, its “future” does not depend on “past”, this process is called Markov process [19]. The mathematical concept is as follows:

Definition 4.1: It is supposed that the state space of random process \{X(t), t \in T\} is \( I \), if time \( t \) has any value of \( n \), that is \( t_{1} < t_{2} < \cdots < t_{n} \), \( n \geq 3 \), \( t_{i} \in T \), under the conditions of \( X(t_{i}) = x_{i}, x_{i} \in I, i = 1,2,\ldots,n-1 \), the condition
distribution function of $X(t_n)$ equal exactly the condition distribution function of $X(t_{n-1})$ under the condition of $X(t_{n-1}) = x_{n-1}$, that is,

$$P\{X(t_n) \leq x_n | X(t_i) = x_i, X(t_{i+1}) = x_{i+1}, ..., X(t_{n-1}) = x_{n-1}\} =$$

$$P\{X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}\}$$  \hspace{1cm} (17)

That we call the process $\{X(t), t \in T\}$ have properties of Markov, or the process is called the Markov process, which both time and state are discrete is called Markov chain. The step of DREM is as follows:

**Step 1:** Evaluation rating

To evaluate recommendation ability of nodes, at first, this article builds evaluation rating, and represents the level with evaluation indexing set $V\{v_1, v_2, v_3, v_4, v_5\}$ of recommendation ability.

**Step 2:** Rating and the proportion of the number of rating

To count the number of different levels of evaluating recommendation nodes, $M_i$ refers to the number of rate in level $i$, $S_i$ shows the proportion of the number of rate in present level, formula is as following:

$$\sum_{i=1}^{5} M_i = N, S_i = \frac{M_i}{N}$$  \hspace{1cm} (18)

**Step 3:** State with the state transition

This article uses the vector $S_u^t = (S_u^1, S_u^2, ..., S_u^5)$ to express the evaluation status of recommended ability of node $w$, referred to as the state. Parameter $t$ ($t$ is a discrete magnitude) represents time, so that the $k$-th evaluation status of node $w$ can be shown as $S_u^k$. When $t$ changes, the evaluation status of $w$ will change with its behavior change, and the state transition will occurred.

**Step 4:** Transition probability matrix

If $S_u^k$ is evaluation status of $w$ in the $k$-th time, the previous evaluation state vector is $S_u^{k-1}$, as all known, $S_u^{k-1} \times P_u = S_u^{k}$ by the Kolmogorov equation[20], in which, $P_u$ is one step transition probability matrix.

$$P_u = \begin{pmatrix} P_{i1} & P_{i2} & \cdots & P_{in} \\ P_{11} & P_{12} & \cdots & P_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{mn} \end{pmatrix}$$  \hspace{1cm} (19)

In which, $P_{ij}$ expresses the transition probability from the level “$i$” to the level “$j$”, and $0 \leq P_{ij} \leq 1$ $(i,j = 1, 2, \cdots, n), \sum_{j=1}^{n} P_{ij} = 1, (i = 1, 2, \cdots, n)$.

Because Markov process has the ergodicity [21], when $t$ tends to $\infty$,  \( S_u^{t-1} = S_u^t = S_u^* \). As long as steady-state $S_u = (S_u^1, S_u^2, ..., S_u^5)$ is calculated, we can calculate the degree of improvement of node $w$ by using $S_u^*$. $S_u^*$ obtained from equation (18):

$$\begin{align*} S_u 	imes P_u &= S_u^* \\ \sum_{i=1}^{5} S_i &= 1 \end{align*}$$  \hspace{1cm} (20)

If $x_1, x_2, ..., x_5$ are given, the progress results can be quantified, and we can obtain scores of recommended ability, and it shows as $F_u^*$,

$$F_u^* = \sum_{i=1}^{5} S_i X_i$$  \hspace{1cm} (21)

Though we obtain the value of $F_u$, $F_u$ does not represent the actual recommendation ability of the node, and just shows that the degree of improvement of node during this time. If the result of $F_u$ is big, only indicates the degree of improvement of node is the bigger to compare with previous one.

**Example 4.1.** If there are recommended nodes $w_1, w_2$, and the evaluation status is divided into 5 levels (high degree of confidence, trust, basic trust, distrust, a high degree of distrust), shown as (HT, T, BT, DT, HDT), the number and transition of state of nodes $w_1, w_2$ are shown as Table I and Table II.

<table>
<thead>
<tr>
<th>Grade</th>
<th>HT</th>
<th>T</th>
<th>BT</th>
<th>DT</th>
<th>HDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HDT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of people</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>30</td>
</tr>
</tbody>
</table>

Table I. Evaluation of Node $w_1$.
and to select we can get steady state of $S_0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix}$. The others are same argument.

The actual status of the last evaluation:

$S_1 = \begin{bmatrix} 5/30, 21/30, 4/30, 0, 0 \\ 2/30, 26/30, 2/30, 0, 0 \\ 6/30, 21/30, 3/30, 0, 0 \\ 4/30, 25/30, 1/30, 0, 0 \\ 2/15, 13/15, 0, 0, 0 \end{bmatrix}$

A transition probability $P_n$ and $P_o$ are as follows:

$$P_n = \begin{bmatrix} 4/5 & 1/5 & 0 & 0 & 0 \\ 2/21 & 18/21 & 1/21 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$P_o = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/13 & 12/13 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

From $(I - P_n)^T S_{w1} = 0$ we can get steady state of $w_1$ and $w_2$:

$$S_{w1} = \begin{bmatrix} 1/5 \\ -1/5 \\ -2/21 \\ 0 \\ 0 \end{bmatrix}, S_{w2} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/13 \\ 0 \\ 0 \end{bmatrix}$$

So we can get $S_{w1} = (10/33, 7/11, 2/33, 0, 0)$.

If the specific score of the evaluation rating is $FT = 90$, $T = 80$, $BT = 60$, $NT = 50$, $NFT = 30$. Scores of recommendation ability of the node $w_1$ and $w_2$ are respectively $F_{w1} = 81.8$, $F_{w2} = 81.3$, improvement of the node $w_1$ is the faster than the node $w_2$.

After recommendation service accuracy of each node is calculated from formula (15) $\Delta = \frac{F_{w1}}{F_{w2}} - 1$, and to select recommendation node that recommendation accuracy is higher than others, then modifies the list of excellent recommendation nodes selected according to getting nodes whose the progress speed of recommendation service is faster than others, so it improved a comprehensive evaluation to the quality of recommendation service, and the final we obtain the list of excellent recommendation nodes.

V. SIMULATION EXPERIMENT AND RESULTS ANALYSIS

Our simulation experiment shows that our proposed trust evaluation model can restrain effectively strategic deception of nodes and malicious recommendation of nodes and dynamically adopt to this model.

We simulate the operation of DRTEMBB model through experiment and verifies the results of the experiment by operating the model in a small-scale network. The experiment involves 100 nodes and 100 services. These services are randomly assigned to each node, marking sure that each node is assigned with at least one service, and each simulation is supposed to comprise several cycles, and there must be one interaction at each node within each cycle. The trust value of the initial node is set at 0.5, indicating it is both trustworthy and untrustworthy. The simulation experiment is conducted under Java environment. And Operating Environment is CPU 3.0G, and Memory is 2G.

Experiment 5.1: Analysis aiming at strategic deception of malicious nodes.

This experiment aim at the issue of strategic deception of malicious nodes with time change, and inspects perception of DRTEMBB model to strategic deception of nodes, and analyses the difference of DRTEMBB model with EigenRep model, the normal model to strategic deception. The experiment assumes malicious nodes adopt suddenly cheating service action after malicious nodes have accumulated higher trust through successful interaction in a period of time.

<table>
<thead>
<tr>
<th>Number of people</th>
<th>The evaluation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$HT$</td>
<td>$T$</td>
</tr>
<tr>
<td>Previous evaluation</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>25</td>
</tr>
</tbody>
</table>

(In table I and table II, the meaning of "*" figure represents that there are five evaluators in previous evaluation, and their evaluation results are "HT", but in the current evaluation, evaluation results of the five peoples are respectively four peoples give "HT", one people gives "T". The others are same argument.)
We find when the rate of malicious nodes is set at 40% and cycles is over 50, EigenRep shows transaction success ratio begins decrease, but DRETMBB is stable. As Cycle is increasing, the effect of DRETMBB is always better than EigenRep model. DRETMBB Model demonstrates a very good success ratio with the increase of transaction cycles.

**Experiment 5.4: Analysis directing to malicious recommendation attack**

Figure 7. Impact of Malicious Recommendation Nodes Percentage to Transaction Success Rate

Figure7.shows, with more and more increasing of dishonest node ratio, that trust computing accurateness declines gradually, which leads to the increasing of transaction failure numbers and the decreasing of the success ratio. Since each dishonest recommendation will influence trust value of itself, and will lead to self trust value to be decreased along with dishonest node ratio increasing, the decline magnitude of the success ratio of transaction is the larger. Hassan model has a kind of resistant ability, but because Hassan assumes that recommendation nodes have higher trust value, meanwhile, Hassan model lacks punishment mechanism, and it cannot effectively weaken the influence of dishonest recommendation to trust, and the success ratio of transaction decreases rapidly.

DRETMBB model shows that the dishonest node can be shielded through the other honest nodes of recommendation path, so decline of transaction success ratio is slower than Hassan model. In contrast to Hassan model, DRETMBB model can filter malicious recommendation nodes more effectively and make recommendation information to be more accurate through leading into trade goods concept similarity and dynamic adjustment trust value.

**VI. Conclusion and Further works**

Trust security is an important research domain in e-commerce environment. Though many researchers have proposed a few trust security models, there are also many questions in these trust models. So we propose a dynamic recommendation trust evaluation model, and we propose a novel evaluation method of QoRS in the model. By experiment results we could know our model is effective and good for restraining malicious recommendation and fraud service actions. At the same time, the novel evaluation method of QoRS can be effectively used to
ascertain the improvement degree of recommendation service ability, and evaluation for QoRS is the accuracy and objective through integrating recommendation success ratio and improvement degree of recommendation service ability, and the method solves effectively a problem that recommendation of the higher trust node is the better than the others, at the same time, and avoids to arise recommendation oligarch. In addition, we propose recommendation nodes bidding method, bidding method can encourage effectively recommendation nodes to the correct recommendation actively.

Our future work will analyze the other methods for recommendation evaluation model such as multi-attributes analysis evaluation method, and fuzzy mathematic analysis for weight of multi-attributes.

ACKNOWLEDGMENT

This work was supported by a grant from National Science and Technology Major Project of the Ministry of Science and Technology of China (No.2012ZX03002001-004) and the National Natural Science Foundation of China (No. 60873071, No.61172090). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[17] J Sabater,C Sierra.REGRET:A reputation model for gregarious societies[A].Fourth Workshop on Deception Fraud and Trust in Agent Societies[C],2001.61-70

Gang Wang, born in 1974, Ph.D.candidate, His current research interests include mainly Trust security, Service computing, Internet of things and cloud computing.

XiaoLin Gui, born in 1966, Ph.D., Professor, Ph.D. supervisor. The chair professor of the Department of Computer Science, Xi’an Jiaotong University, His research interests include Service computing, Internet of things, dynamic trust management and cloud computing.
Block-Based Parallel Intra Prediction Scheme for HEVC

Jie Jiang
Institute of Intelligence Control and Image Engineering, Xidian University, Xi’an, China
Email: supergirl.jj@gmail.com

Baolong Guo and Wei Mo
Institute of Intelligence Control and Image Engineering, Xidian University, Xi’an, China
Email: {blguo, wmo}@xidian.edu.cn

Kefeng Fan
Guilin University of Electronic Technology, Guilin 541004, China
Email: fankf@cesi.ac.cn

Abstract — Advanced video coding standards have become widely deployed in numerous products, such as multimedia service, broadcasting, mobile television, video conferences, surveillance systems and so on. New compression techniques are gradually included in video coding standards so that a 50% compression rate reduction is achievable every ten years. However, dramatically increased computational complexity is one of the many problems brought by the trend. With recent advancement of VLSI (the Very Large Scale Integration) semiconductor technology contributing to the emerging digital multimedia word, this paper intends to investigate efficient parallel architecture for the emerging high efficiency video coding (HEVC) standard to speed up the intra coding process, without any prediction modes ignored. Parallelism is achieved by limiting the reference pixels of the 4 × 4 subblocks, allowing the subblocks to use different direction modes to predict the residuals. Experimental implementations of the proposed algorithm are demonstrated by using a set of video test sequences that are widely used and freely available. The results show that the proposed algorithm can achieve a satisfying intra parallelism without any significant performance lose.

Index Terms — HEVC, intra coding, parallel architecture, multiple directions.

I. INTRODUCTION

Continuous emergence of video coding standards and the growth in development and implementation technology for them have undoubtedly created a completely new world of multimedia. So far, contributions to video coding technology have mainly focused on improving coding efficiency. The challenges remain: not only to find efficient coding algorithms which require high performance but also to speed up the coding process.

The ongoing video coding standard, High Efficiency Video Coding (HEVC) [1], is getting more attention due to its high compression efficiency. However, the computational complexity of HEVC would be 2-10 times higher than its counterpart, which is considered an obstacle to implement it in real-time. Therefore, many research works focus on how to reduce the computational complexity.

The purpose of these works is to design and evaluate the performance of new methods to reduce encoder complexity, while keeping the quality of reconstructed video sequences for intra coding. The works generally fall into two categories.

1. Fast mode decision approaches with early termination using adaptive thresholds or optimized Lagrangian rate distortion optimization (RDO) function [2-4].
2. Parallel architectures to speed up the intra prediction process [5-14].

With recent advancement of VLSI (the Very Large Scale Integration) semiconductor technology contributing to the emerging digital multimedia word, research on parallel architectures gets more attention. In this paper, we focus on the second case, and present a block based parallel architecture to speed up the intra prediction for HEVC.

The remaining parts of this paper are organized as follows: Section II reviews the state of art within the field of parallel architectures. Section III introduces the spatial prediction in HEVC. Section IV presents the proposed scheme, including 2X parallel intra prediction and its expansion to 4X parallelism. Experimental results are presented within Section V. Finally, we conclude this paper in section VI.

II. RELATED WORK

The main image and intra frame of video compression extensively adopts the block-based structure from prediction and transform to entropy coding, where the coding of one block is dependent on the availability of its left, upper-left, and upper-right blocks. Such a highly dependent structure is not quite suitable for parallelization, especially for ASIC (Application Specific Integrated Circuit) solutions. Even so, when dual-core and quad-core computers are available, there are still many efforts on parallelizing the encoding and decoding from different aspects, as described below.

1. GOP (Group of Pictures) approaches: Barbosa [5] and Vander [6] propose to partition a sequence into some GOPs. The correlation between GOPs is low, and it can not only limit error propagation, but also support parallel coding processing. However, it needs to get the data of all the pictures in a GOP before parallelism. When the GOP has too
many pictures, it will lead to serious delay, which is not convenient for the real-time video coding applications.

2. Frame approaches: Chen [7] proposes to realize the parallelism coding at the frame level. This approaches is limited to the correlation between frames, therefore, the speedup cannot be linearly increased corresponding to the number of process cores.

3. Pipeline approaches: Gulati [8] and Klaus [9] propose to organize the prediction, transform, and entropy coding of macroblocks (MBs) as a pipeline and assign them to multiple cores for parallel computing. This class of approaches can achieve limited parallelisms if workloads are unbalanced at different cores. Two times speedup is reported for high definition sequences on general-purpose quad-core computers in [9].

4. Slice partitioning (SP) approaches: Rodriguez [10] proposes to partition an image into some regions that are referred to as slices. The coding of slices can be carried out independently by different cores. They provide good parallelism but would result in a significant loss on coding performance if there are too many slices.

5. MB-reordering (MR) approaches: Despite these efforts on parallelizing existing coding algorithms, due to strong dependency among blocks in intra prediction and the filtering process of top/left reconstructed samples, the intra luma prediction process of small blocks is a challenge for parallelization. The original process handles blocks in serial, which is not efficient. Efficient architectures have been reported in [11-14], which propose to process MBs in the wave front order so that MBs in each diagonal line can be coded concurrently when neighboring MBs are available. Owing to the fine-granularity parallelism at the MB level, these approaches can achieve good parallelism and are more widely used at present.

Huang's work [11] has bubbles between Intra 4 × 4 predictions because of the low throughput of reconstruction process so that the prediction has to wait for the completion of reconstruction. Lee's work [12] perfectly pipelines the intra prediction and reconstruction process; however, it requires that both intra prediction and reconstruction have exact equal processing cycles. It also reduces some prediction modes in some blocks in order to enforce pipelining; hence, the video quality is degraded. Jin's work [13] proposes both partially and fully pipelined architectures for intra 4 × 4 prediction and has the same drawback as the approach in [12]. Moreover, the architectures add dependency graph process in order to improve gains; however, this increases hardware overhead. It takes 25 cycles to process each block, which is too long for high throughput reconstruction. Similar to Huang's work, Suh's work [14] proposes an efficient parallel architecture followed by a redundancy reduction algorithm to speed up the intra 4 × 4 prediction. However, the approaches above all have drawbacks either with the pipelining architecture or in compression gains.

In this paper, we propose a block-based intra prediction parallel algorithm to solve the problem above.

III. Spatial Prediction in HEVC

ISO-IEC/MPEG and ITU-T/VCEG recently formed the joint collaborative team on video coding (JCT-VC). The JCT-VC aims to develop the next-generation video coding standard, called high efficiency video coding (HEVC). A joint proposal to the high efficiency video coding standardization effort have been partially adopted into the HEVC Test Model (HM). The major improvements for intra coding come from two tools described below.

A. Intra Prediction

The nine intra prediction modes supported in H.264/AVC with different directionalities is not flexible enough to represent complex structures or image segments with different directionalities. HEVC extends the set of directional prediction modes of H.264/AVC, providing increased flexibility and more accurate predictions for the sample values. The increased prediction accuracy provides significant reductions in residual energy of the intra coded blocks and improvements in coding efficiency [15].

It can support from 16 to 34 prediction directions for different prediction size, and the prediction can achieve 1/8 pixel accuracy. Table I shows a 64 × 64 prediction unit contains different subblock size and its corresponding prediction modes. The 34 directions of the intra prediction modes are illustrated in Fig. 1.

<table>
<thead>
<tr>
<th>Amount</th>
<th>Size</th>
<th>Modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>4x4</td>
<td>17</td>
</tr>
<tr>
<td>64</td>
<td>8x8</td>
<td>34</td>
</tr>
<tr>
<td>16</td>
<td>16x16</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>32x32</td>
<td>34</td>
</tr>
<tr>
<td>1</td>
<td>64x64</td>
<td>5</td>
</tr>
</tbody>
</table>

It improves the coding efficiency significantly through permitting block prediction in an arbitrary direction by indicating the prediction angle.
B. Quadtree-based coding structure

Coding efficiency can be significantly improved by utilizing macroblock structures with sizes larger than $16 \times 16$ pixels, especially at sequences with high resolutions [16]. To achieve a more flexible coding scheme, HEVC utilizes a quadtree-based coding structure [17] with support for macroblocks of size $64 \times 64$, $32 \times 32$, $16 \times 16$, $8 \times 8$, and $4 \times 4$ pixels. HEVC separately defines three block concepts: coding unit (CU), prediction unit (PU) and transform unit (TU). After the size of largest coding unit (LCU) and the hierarchical depth of CU have been defined, the overall structure of codec is characterized by the various sizes of CU, PU and TU in a recursive manner. This allows the codec to be readily adapted for various kinds of content, applications, or devices that have different capabilities.

The CU splitting process is described in Fig. 2, using a Split flag to ensure whether the current CU needs to split into a smaller size or not.

![Figure 2. The CU splitting process](image)

The coding efficiency improvements are more visible at higher resolutions, where bit rate reductions reach around 50% and 35% for the low delay and random access experiments, respectively.

C. Observation and Motivation

Intra prediction is currently achieved by partitioning a largest coding unit (LCU) into one or more blocks through a recursive splitting process. Each block is predicted spatially and subsequently refined, and the prediction is performed sequentially using neighboring reconstructed blocks. The prediction process requires that the causal neighbors of the current block must be completely reconstructed before processing the current block. It results in a set of serial dependencies, and these serial dependencies result in significant complexity for both the encoder and decoder processes.

![Figure 3. An $8 \times 8$ block with four $4 \times 4$ blocks](image)
reconstructing the first set of blocks, we proceed to predict the second set of blocks. As mentioned before, these blocks are predicted in parallel and use the reconstructed pixel values from blocks in the first partition and the left coding unit’s boundaries, as showed in Fig. 4 D.

In the proposed scheme, none of the prediction modes are ignored. That means it support all the prediction directions mentioned in Part A, Section III. Therefore, it can realize the parallelism without significant performance loss.

Figure 4. The blocks’ reference pixels

By employing the partitioning strategy, we can achieve a direct increase in parallelism. This parallelism is shown graphically in Fig. 5. The 4 × 4 intra prediction within an 8 × 8 coding unit requires 4 sequential steps while our partitioning approach results in only two sequential steps. This is an increase in parallelism by a factor of 2X. In general, the increase in parallelism is N/2, where N is the number of blocks within the macro-block.

Figure 5. Processing order for the 4×4 blocks within an 8×8 intra predicted unit

B. 4X parallel intra prediction

The method can also be extended the parallelism to be 4X. In this extension, all of four blocks are treated as the first set blocks. They are restricted to have the same prediction mode, and are simply predicted using their 8 × 8 block neighbors, as shown in Fig. 6. However, each 4 × 4 block can have its own predict mode, residual and transform. In this case, the distances between the referenced pixels and predicting pixels increase, therefore it will lead to a bit rate increase, as will be seen in the results later.

Figure 6. The blocks’ reference pixels for 4X parallelism

V. EXPERIMENTAL RESULTS

The parallel intra prediction technology has been integrated into the HEVC reference software HM3.0 [18] and tested the recommended configuration on 18 sequences listed in Table II.

The test sequences include two 4K sequences, five 1080p sequences, four WVGA sequences, four WQVGA sequences, and three 720P sequences, which are widely used in various applications.

We use two common test conditions described in [19]. Intra (high efficiency condition) and Intra LoCo (low complexity condition).

The common test conditions are set as follows.
1) 10 seconds length are encoded as intra frames for each sequence.
2) TheQP is set at 22, 27, 32, and 37.
3) In-loop deblocking filter and RDO are enabled.

Table II: Comparison of the performance for 1DDCT and 2DDCT

<table>
<thead>
<tr>
<th>Class A 1080p</th>
<th>Class B 1080p</th>
<th>Class C WVGA</th>
<th>Class D WVQGA</th>
<th>Class E 720P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic</td>
<td>Kimono</td>
<td>BasketballDrill</td>
<td>BasketballPass</td>
<td>Vidy1</td>
</tr>
<tr>
<td>PeopleOn Street</td>
<td>ParkScene</td>
<td>BQMall</td>
<td>BQSqure</td>
<td>Vidy3</td>
</tr>
<tr>
<td>Cactus</td>
<td>PartyScene</td>
<td>BlowingBubbles</td>
<td>Vidy4</td>
<td></td>
</tr>
<tr>
<td>BasketballDrive</td>
<td>RaceScene</td>
<td>RaceHorses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BQTerrace</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table III below shows the summary results of 2X parallel intra prediction vs. HM3.0. Table IV shows summary results of 4X parallel intra prediction vs. HM3.0. Y BD-rate means the BD-rate increase for luma component, while U BD-rate and V BD-rate for the chroma component. Compare Table III with Table IV, we can find that the coding performance of 4X parallelism is not as good as 2X parallelism. For the 4X parallelism case, the distances between the referenced pixels and predicting pixels increase, and that reduces the prediction accuracy, which results in the performance lose.
TABLE III
RD results of 2X parallel intra prediction with vs. HM3.0

<table>
<thead>
<tr>
<th></th>
<th>Intra LoCo</th>
<th></th>
<th>Intra LoCo</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y BD-rate%</td>
<td>U BD-rate%</td>
<td>V BD-rate%</td>
<td>Y BD-rate%</td>
</tr>
<tr>
<td>Class A</td>
<td>0.6</td>
<td>0.1</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Class B</td>
<td>0.7</td>
<td>0.6</td>
<td>-0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Class C</td>
<td>1.5</td>
<td>1.1</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Class D</td>
<td>1.8</td>
<td>1.4</td>
<td>0.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Class E</td>
<td>0.7</td>
<td>1.2</td>
<td>-0.5</td>
<td>1</td>
</tr>
<tr>
<td>All</td>
<td>1.1</td>
<td>0.8</td>
<td>0.2</td>
<td>1.1</td>
</tr>
</tbody>
</table>

TABLE IV
RD results of 4X parallel intra prediction with vs. HM3.0

<table>
<thead>
<tr>
<th></th>
<th>Intra LoCo</th>
<th></th>
<th>Intra LoCo</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y BD-rate%</td>
<td>U BD-rate%</td>
<td>V BD-rate%</td>
<td>Y BD-rate%</td>
</tr>
<tr>
<td>Class A</td>
<td>1.1</td>
<td>-0.6</td>
<td>-0.7</td>
<td>1.1</td>
</tr>
<tr>
<td>Class B</td>
<td>1.2</td>
<td>0.6</td>
<td>0</td>
<td>1.3</td>
</tr>
<tr>
<td>Class C</td>
<td>2.7</td>
<td>1.9</td>
<td>1</td>
<td>3.3</td>
</tr>
<tr>
<td>Class D</td>
<td>3.5</td>
<td>1.4</td>
<td>1.3</td>
<td>3.5</td>
</tr>
<tr>
<td>Class E</td>
<td>1.6</td>
<td>1</td>
<td>0.7</td>
<td>2.1</td>
</tr>
<tr>
<td>All</td>
<td>2</td>
<td>0.9</td>
<td>0.4</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Due to the various special applications of intra coding frame such as random access point and refresh synchronous frame, intra coding is an important part of video coding. For the intra coding blocks are predicated dependent on the availability of the adjacent blocks in the former rows and columns, so the parallel coding or decoding process cannot be realized among the intra coding blocks, which increases the difficulties for the implementation of real-time coding in high definition and ultra high definition applications. Therefore, how to increase the capacity of parallel computing effectively has considerable value.

In this paper, a parallel prediction scheme for HEVC is proposed. The parallel prediction unit supports the parallelization of intra-coded blocks within the current coding unit in a two-step parallel process. Parallelism is achieved by limiting the reference pixels of the 4 × 4 subblocks, without any prediction modes ignored. Experimental results show that the increased parallelization comes with small losses in coding efficiency. For example, about 1% for HD sequences of 2X parallelism. We assert that this loss is negligible and well justified by the parallelization capability.

VI. CONCLUSIONS

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 60802077; No. 61003196; No. 61172053)

REFERENCES


Jie Jiang received the B.S. and M.S. degree both in signal and information processing from Xidian University, Xi’an, China, in 2005 and 2008, respectively. She is currently working toward the Ph.D. degree with the Institute of Intelligent Control and Image Engineering (ICIE), Xidian University. Her research interests include video coding and communication.

Baolong Guo received his B.S., M. S. and Ph.D. degrees from Xidian University in 1984, 1988 and 1995, respectively, all in communication and electronic system. From 1998 to 1999, he was a visiting scientist at Doshisha University, Japan. He is currently a full professor with the Institute of Intelligent Control and Image Engineering (ICIE) at Xidian University. His research interests include intelligent information processing, image processing and communication.

Wei Mo was born in Guangxi Province, China, in 1956. He is a professor, Ph.D. supervisor at Xidian University. His research interests include multimedia coding and intelligent information processing.

Fan Kefeng completed his Postdoctoral Research in the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China. His research interests include 3D TV, Smart TV, DRM, digital interface, et al.
Optimized LSB Matching Steganography Based on Fisher Information

Yi-feng Sun
Zhengzhou Information Science and Technology Institute, Zhengzhou, China
Email: yfsun001@163.com

Dan-mei Niu
Electronic Information Engineering College, Henan University of Science and Technology, Luoyang, China
Email: niudanmei@163.com

Guang-ming Tang and Zhan-zhan Gao
Zhengzhou Information Science and Technology Institute, Zhengzhou, China
Email: tgm1983@sina.com, gaozhandyx@126.com

Abstract—This paper proposes an optimized LSB matching steganography based on Fisher Information. The embedding algorithm is designed to solve the optimization problem, in which Fisher information is the objective function and embedding transferring probabilities are variables to be optimized. Fisher information is the quadratic function of the embedding transferring probabilities, and the coefficients of quadratic term are determined by the joint probability distribution of cover elements. By modeling the groups of elements in a cover image as Gaussian mixture model, the joint probability distribution of cover elements for each cover image is obtained by estimating the parameters of Gaussian mixture distribution. For each sub-Gaussian distribution in Gaussian mixture distribution, the quadratic term coefficients of Fisher information are calculated, and the optimized embedding transferring probabilities are solved by quadratic programming. By maximum posteriori probability principle, cover pixels are classified as the categories corresponding to sub-Gaussian distributions. At last, in order to embed message bits, pixels chose to add or subtract one according to the optimized transferring probabilities of the category. The experiments show that the security performance of this new algorithm is better than the existing LSB matching.

Index Terms—steganography, security, information hiding, quadratic programming, Fisher information

I. INTRODUCTION

The aim of steganography is to hide secret message imperceptibly into a cover, so that the presence of hidden data cannot be diagnosed. But steganography faces the threaten of steganalysis. Steganalysis aims to expose the presence of hidden data. How to improve the security of steganography is an important problem.

In Least Significant Bit (LSB) replacement algorithm, the LSBs of cover elements are replaced with message bits. Some structural asymmetry (never decreasing even pixels and increasing odd pixels when hiding message bit) is introduced. It is easy to detect the existence of hidden message. A trivial modification of LSB replacement is LSB matching, which randomly increases or decreases pixel values by one to match the LSBs with the message bits. LSB matching is much harder to detect than LSB replacement algorithm.

The embedding in LSB matching is similar to adding an independent and identically distributed (IID) noise sequence independent of cover. It is well known that values of neighboring pixels in natural images are not independent. Further more, there are complex dependences in the noise component of neighboring pixels [1]. These dependences are violated by LSB matching. Many steganalysis methods utilized this fact [1][2]. The LSB matching also needs to be improved.

There are two kinds of approaches to improve steganography. The first approach is to preserve a chosen cover model in steganographic methods, such as model-based (MB) steganography[3][4], OutGuess[5], the statistical restoration based steganographic algorithms [6][7], which preserve the first and second order statistics. Another strategy is to minimize a heuristically-defined embedding distortion. The steganography using tri-way pixel-value differencing [8] reduces the quality distortion of stego-image. Matrix embedding methods [9][10][11] minimize the change number of cover elements by linear codes, and the change number is predefined as embedding distortion. Many other syndrome codes steganographic methods, such as steganography using wet paper codes[12], steganographic algorithms using syndrome trellis codes [13][14][15][16], minimize the more complex embedding distortion. According to experiments, the above steganographic methods have the better performance.

In this paper, we propose an optimized LSB matching steganography based on Fisher information. It can
demonstrate the security improvement in the sense of Kullback Leibler (KL) divergence. Firstly, we introduce the relation between KL divergence and Fisher information, and explain why Fisher information can be used to improve steganographic security. Then we present an embedding optimization framework based on Fisher information. In the framework, an embedding algorithm is designed to solve the optimization problem whose objective is minimizing the Fisher Information. The optimized LSB matching algorithm is an instance of the framework. We assume that the groups of pixels in the cover image submit Gaussian mixture distributions. After obtaining the Gaussian mixture distribution of pixel group for each cover image, the optimized embedding transferring probabilities are solved by quadratic programming for each sub-Gaussian distribution. Cover pixels are classified as the categories corresponding to sub-Gaussian distributions. The embedding is operated by the optimized embedding transferring probabilities. The experimental results show that the optimized LSB matching steganography is better than the existing LSB matching in the aspect of security performance.

II. STEGANOGRAPHIC SECURITY AND FISHER INFORMATION

KL divergence between cover and stego distributions can be used to measure the steganographic security [17]. Let the distribution of cover objects be \( P_0 \), and the distribution of stego objects be \( P_s \), where the parameter \( \lambda \) denotes the relative payload, and \( 0 \leq \lambda \leq 1 \). The KL divergence between \( P_0 \) and \( P_s \) is denoted as \( D(P_0 \parallel P_s) \). The smaller the \( D(P_0 \parallel P_s) \) is, the more secure the steganography will be. If \( D(P_0 \parallel P_s) \) is equal to 0, stego-system is perfectly secure.

The physical interpretation of cover objects is not discussed in [17]. We think that each statistic from a cover can represent the cover object. Here the elements sequence of cover, such as the sequence of the pixel value, or the sequence of Discrete Cosine Transform (DCT) coefficients, is selected as the presentation of the cover object. The elements sequence is denote by \( C = (C_1, C_2, \ldots, C_N) \), where \( N \) is the number of the elements that can be used to embed secret information in a multimedia cover. In the following, we use \( P_0^{(N)} \) and \( P_s^{(N)} \) instead of \( P_0 \) and \( P_s \). It emphasizes that \( P_0^{(N)} \) and \( P_s^{(N)} \) is \( N \)-dimensional joint probability distribution.

Assuming that \( P_0^{(N)} \) is known and embedding function is fixed, KL divergence \( D(P_0^{(N)} \parallel P_s^{(N)}) \) is the function of \( \lambda \). KL divergence can also be denoted as \( d_\lambda(\lambda) \). The second derivative of \( d_\lambda(\lambda) \) to \( \lambda \) (when \( \lambda = 0 \)) is known as steganographic Fisher Information (FI) [18]. It is denoted by \( d_\lambda^2(0) \). When \( \lambda \to 0 \),

\[
D(P_0^{(N)} \parallel P_s^{(N)}) \approx \frac{1}{2} d_\lambda^2(0) \cdot \lambda^2.
\]  

When relative payload \( \lambda \) is fixed, Fisher Information is smaller, and KL divergence will be smaller. Thus Fisher Information can be used for evaluating steganographic security.

Assuming that embedding operations are mutually independent (MI) [19], Fisher Information can be calculated. The probability of cover element being \( x \) and stego element being \( y \) is denoted by the conditional probability \( P(S_n = y | C_n = x) = b_{ny}, x \in \chi, y \in \chi, \) where \( \chi \) is the set of cover and stego elements. The matrix \( B = (b_{ny}), \) also named as embedding matrix, corresponds to an embedding method. The probability \( b_{ny} \) is also called as the embedding transferring probability. For embedding matrix \( B \), Fisher information is calculated [20]:

\[
d^*_a(0) = \sum_{y \in \chi} \frac{A(y)^2}{P_0^{(N)}(C = y)} - N^2,
\]

\[
A(y) = \sum_{u \in \chi} \sum_{y \in \chi} [P_0^{(N)}(C = y \mid u, y) - b_{uy}].
\]

Here \( y = (y_1, y_2, \ldots, y_N) \in \chi^N \) denotes the value of \( C = (C_1, C_2, \ldots, C_N) \), where \( \chi^N = \chi \times \chi \times \cdots \times \chi \). Here \( y \mid u, y \) denotes the sequence \( (y_1, \ldots, y_i, u, y_{i+1}, \ldots, y_N) \).

It means that \( u \) replaces the item \( y_i \) in \( y \).

The computational complexity of Fisher information is decided by \( N \). We call \( d^*_a(0) \) \( N \)-dimensional Fisher Information. We often simplify the model of the covers in order to calculate KL divergence. For example, the elements of cover objects are IID. But it ignores the correlation among the elements of cover objects. Hence, we suppose that the cover is composed of IID element groups. The elements in a group are correlated, and the number of elements in a group is \( n \). If a group is composed of two elements, we can take the impact of second-order correlation into account. If it includes more elements, we can analyze the influence of "higher order" correlation. For cover \( C = (C_1, C_2, \ldots, C_N) \), starting from the first element \( C_1 \), every \( n \) adjacent elements are divided into a group, whose \( n \)-dimensional joint distribution is denoted by \( P_0^{(n)} \). If \( P_0^{(n)} \) replaces \( P_0^{(N)} \), and \( n \) replaces \( N \) in (2) and (3), we can obtain the calculation formula about \( n \)-dimensional Fisher information corresponding to the element groups. The \( n \)-dimensional Fisher information is denoted by \( d^*_a(0) \). In the following, we focus on decreasing the value of \( d^*_a(0) \) to improve the security of embedding method.

III. THE EMBEDDING OPTIMIZATION FRAMEWORK BASED ON FISHER INFORMATION

In order to improve the steganographic security, we need to find the embedding algorithm whose \( B \) matrix corresponds to the small Fisher information. In general,
the embedding matrix \( \mathbf{B} = (b_{xy})_{x,y} \) is usually constant for each cover. For example, in LSB matching algorithm, if the message bit does not match the LSB of the cover pixel, one is randomly either added or subtracted from the value of the cover pixel. That is to say \( b_{xy(0)} = b_{xy(-1)} = 0.25 \) and \( b_{xy} = 0.5 \) for each cover. Here we think the probability \( b_{xy} \) is variable, and we adjust \( b_{xy} \) to minimize \( d^*_{xy}(0) \) for each cover, assuming that \( P_{0}(\mathbf{y}) \) can be obtained according to one cover. So \( \mathbf{B} = (b_{xy})_{x,y} \) can be adapted according to the cover statistical properties, and we can get optimized embedding algorithm.

Here \( d^*_{xy}(0) \) is determined by the \( n \)-dimensional joint probability distribution \( P_{0}(\mathbf{y}) \) and embedding matrix \( \mathbf{B} = (b_{xy})_{x,y} \). If \( b_{xy}, x \in \chi, y \in \chi \) is viewed as variables, Fisher information is the function of \( b_{xy} \). To improve the steganographic security, we need solve the optimization problem in which the Fisher information \( (0) \) is the objective function. Here \( n \) is constant, and the optimization problem is

\[
\mathbf{B}_{\text{optimized}} = \arg \min_{\mathbf{B}(b_{xy})} \sum_{y \in \chi} \frac{A(y)^2}{P_{0}(\mathbf{y}) \text{E} \{\mathbf{y} = \mathbf{y}\}},
\]

with the constraints
\[
\sum_{y \in \chi} b_{xy} = 1, x \in \chi
\]
\[
b_{xy} \geq 0
\]

We expand the formula as follows:

\[
\sum_{y \in \chi} \frac{A(y)^2}{P_{0}(\mathbf{y}) \text{E} \{\mathbf{y} = \mathbf{y}\}} = \sum_{\mathbf{y} \in \chi} \sum_{u \in \chi} \sum_{v \in \chi} P_{0}(\mathbf{y}) P_{u}(\mathbf{v}) P_{v}(\mathbf{y}) h_{uyv},
\]

\[
= \sum_{\mathbf{y} \in \chi} \sum_{u \in \chi} \sum_{v \in \chi} \sum_{\mathbf{z} \in \chi} P_{0}(\mathbf{y}) P_{u}(\mathbf{v}) P_{v}(\mathbf{y}) h_{uyv} \left\{ \begin{array}{l}
\sum_{j=1}^{K} \sum_{a_{ij} \in \chi} P_{a_{ij}}(\mathbf{y}) P_{a_{ij}}(\mathbf{v}) P_{a_{ij}}(\mathbf{y}) h_{uyv} \\
\sum_{j=1}^{K} \sum_{a_{ij} \in \chi} P_{a_{ij}}(\mathbf{y}) P_{a_{ij}}(\mathbf{v}) P_{a_{ij}}(\mathbf{y}) h_{uyv} \\
\sum_{j=1}^{K} \sum_{a_{ij} \in \chi} P_{a_{ij}}(\mathbf{y}) P_{a_{ij}}(\mathbf{v}) P_{a_{ij}}(\mathbf{y}) h_{uyv}
\end{array} \right\} h_{uyv},
\]

\[
+ \sum_{\mathbf{y} \in \chi} \sum_{u \in \chi} \sum_{v \in \chi} \sum_{\mathbf{z} \in \chi} P_{0}(\mathbf{y}) P_{u}(\mathbf{v}) P_{v}(\mathbf{y}) h_{uyv} \left\{ \begin{array}{l}
\sum_{j=1}^{K} \sum_{a_{ij} \in \chi} P_{a_{ij}}(\mathbf{y}) P_{a_{ij}}(\mathbf{v}) P_{a_{ij}}(\mathbf{y}) h_{uyv} \\
\sum_{j=1}^{K} \sum_{a_{ij} \in \chi} P_{a_{ij}}(\mathbf{y}) P_{a_{ij}}(\mathbf{v}) P_{a_{ij}}(\mathbf{y}) h_{uyv} \\
\sum_{j=1}^{K} \sum_{a_{ij} \in \chi} P_{a_{ij}}(\mathbf{y}) P_{a_{ij}}(\mathbf{v}) P_{a_{ij}}(\mathbf{y}) h_{uyv}
\end{array} \right\} h_{uyv}.
\]

Here \( y[a/y, y/b, y/c] \) denotes the sequence

\[
(y_{1}, \cdots, y_{n}, a_{1}, y_{1}, \cdots, y_{n}, b_{1}, y_{1}, \cdots, y_{n}).
\]

From (6), we can find that Fisher information \( d^*_{xy}(0) \) is a quadratic function about the probability \( b_{xy} \). Searching \( b_{xy} \) which minimizes Fisher information is a quadratic programming problem.
pixels should add or subtract 1 according to the transferring probabilities indicated by the optimized embedding matrix corresponding to each category.

\[ i = 1, \ldots, N_1 - 1 \quad \text{and} \quad j = 1, \ldots, N_2 - 1, \]

there are \((N_1 - 1)(N_2 - 1)\) samples in total; at last, every two adjacent pixels in the vice diagonal direction \( \chi \) also constitute \((N_1 - 1)(N_2 - 1)\) samples. With the above \(4(N_1 - 1)(N_2 - 1) + (N_1 - 1) + (N_2 - 1)\) samples, we use the MDL algorithm of Purdue University [21] to estimate the parameters of GMM. The MDL algorithm is an extension of Expectation Maximum (EM) algorithm. Please refer to the literature [22].

Note that GMM is the continuous distribution. We need the discrete distribution in optimization problem (4). In order to reduce computational complexity, the values of the sub-Gaussian probability density \( p_{0,\alpha}(y) \) replace the discrete values directly.

\[ B = \begin{pmatrix}
0.5 & 0.5 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0.5 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0.5 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0.5 & 0.5 \\
\end{pmatrix} \]
For each sub-Gaussian distribution of GMM, we need to obtain the corresponding optimized LSB matching embedding matrix, denoted as \( \mathbf{B}_k^{\text{optimized}} \), \( k = 1, \cdots, K \). Here the pixel group is made up of two adjacent pixels, and the optimization problem is simplified as

\[
\mathbf{B}_k^{\text{optimized}} = \arg \min_{\mathbf{B}} f_k(\mathbf{B}),
\]

\( f_k(\mathbf{B}) = \sum_{i,j} \sum_{k} \sum_{i,j} \left( \sum_{i,j} \mathbf{P}_{0,k}(y_i = u_i, y_j = b) \mathbf{P}_{1,k}(y_i = a, y_j = v) \right) h_{u} h_{b} \)

\( + \sum_{i,j} \sum_{k} \sum_{i,j} \left( \sum_{i,j} \mathbf{P}_{0,k}(y_i = u_i, y_j = a) \mathbf{P}_{1,k}(y_i = a, y_j = v) \right) h_{a} h_{a} \)

\( + \sum_{i,j} \sum_{k} \sum_{i,j} \left( \sum_{i,j} \mathbf{P}_{0,k}(y_i = u_i, y_j = v) \mathbf{P}_{1,k}(y_i = a, y_j = a) \right) h_{u} h_{a} \).

(10)

The constraints are

\[
\begin{align*}
& b_{x+1} + b_{x} = 0.5, \quad b_{x+1} \geq 0, \quad b_{x} \geq 0, \quad x \in \mathbb{Z}, \quad x \neq 0, x \neq 255 \\
& b_{y} = 0.5, \quad x \in \mathbb{Z} \\
& b_{y} = 0.5, b_{y+256} = 0.5 \\
& \text{others} \quad b_{y} = 0.
\end{align*}
\]

(12)

This optimization problem is a quadratic programming. The standard form of quadratic programming is

\[ \min_{x} \frac{1}{2} x^T H x + c^T x \]

where \( x \) is a row vector, and \( H \) represents the quadratic coefficient matrix. By arranging matrix \( \mathbf{B} \) into a row vector, we can get the standard form. The elements of \( H \) are the corresponding coefficients in (6).

The complexity degree of quadratic programming depends on the element number of \( x = \{x_1, x_2, \cdots, x_z\} \) is a row vector, and \( H \) represents the quadratic coefficient matrix. By arranging matrix \( \mathbf{B} \) into a row vector, we can get the standard form. The elements of \( H \) are the corresponding coefficients in (6).

Here, though the amount of elements in matrix \( \mathbf{B} \) is 256 \( \times \) 256, many matrix elements are 0 in our algorithm because \( b_{y} = 0 \) when \( |x - y| \geq 2 \). If \( b_{y} = 0 \), its corresponding quadratic coefficient has no effect on the quadratic function. There are 766 \( \times \) 766 nonzero elements in \( \mathbf{H} \), which need to be stored in the computer memory. They are \( b_{x+1} \), \( b_{x+1} \), where \( x = 1, 2, \cdots, 254 \), \( b_{x} = 0.5 \), where \( x = 0, 1, \cdots, 255 \), \( b_{y} = 0.5 \) and \( b_{y+256} = 0.5 \). PC can store them. So we use quadratic programming function in Matlab software to solve the optimized transferring probabilities \( b_{x+1} \) and \( b_{x+1} \).

C. Judging the category of Pixels

In the existing LSB matching algorithm, if the message bit does not match the LSB of the cover pixel, one is randomly either added or subtracted from the value of the cover pixel. In our optimized LSB matching algorithm, if the message bit does not match the LSB of the cover pixel, we need judging the category to which the pixel belongs at first. The embedding matrix of the 4th category is \( \mathbf{B}_4^{\text{optimized}} \). We add or subtract 1 by the probability \( b_{x+1} \) or \( b_{x+1} \) in \( \mathbf{B}_4^{\text{optimized}} \).

Now we explain the method of judging the category the pixels \( C_{i,j} \) belonging to, where \( i, j \) are the row and column index of the pixel. Firstly, we constitute pixel group by \( C_{i,j} \) and its adjacent pixels. For example, constructing the pixel group \( (C_{i,j}, C_{i,j+1}) \) with the right horizontal adjacent pixel \( C_{i,j+1} \). Note that if \( C_{i,j} \) is at the right edge of the image, then we can use the left horizontal adjacent pixel \( C_{i,j-1} \) to constitute the pixel group. Then based on the value of the pixel group \( y = (C_{i,j}, C_{i,j+1}) \) or \( y = (C_{i,j-1}, C_{i,j}) \), the posterior probabilities of each sub-Gaussian distribution is calculated as follows:

\[
P(k | y) = \frac{P_{0,k}(C = y) \pi_k}{\sum_{k=1}^{K} P_{0,k}(C = y) \pi_k}, \quad k = 1, 2, \cdots, K.
\]

(13)

The pixel \( C_{i,j} \) is categorized according to maximum posteriori probability principle

\[
k^{\text{optimal}} = \arg \max_{k=1,2,\cdots,K} P_{0,k}(C = y) \pi_k,
\]

(14)

where \( k^{\text{optimal}} \) denotes the category of pixels \( C_{i,j} \) to be determined.

D. Embedding and Extraction Algorithm

The embedding process of the optimized LSB matching steganography is summarized as follows:

Input: cover image \{\( C_{i,j} \)\} which has \( N = N_1 \times N_2 \) pixels, \( M \) bits encrypted secret message, stego-key \( k_i \).

Output: stego image \{\( S_{i,j} \)\}.

Step(i): Model the two-dimensional probability distribution of two spatial adjacent pixels as GMM, and estimate the parameters of GMM according to the pixel group samples from cover image \{\( C_{i,j} \)\}.

Step(ii): For each sub-Gaussian distribution \( P_{0,k}(y) \) of GMM, work out the matrix \( \mathbf{H} \) by calculating the coefficients of nonzero \( b_{x+1} \) and \( b_{x+1} \) in objective function \( f_k(\mathbf{B}) \) according to (11), then solve the quadratic programming problem (10), and get the optimized embedding matrix \( \mathbf{B}_k^{\text{optimized}} \) corresponding to each sub-Gaussian distribution.

Step(iii): Calculate the rate of pixels used for embedding, which is denoted as \( \lambda = M / N \). The rate is equal to the relative payload in general sense. According to \( \lambda \) and the stego-key \( k_i \), select the set of image pixels \{\( C_{i,j} \)\} to embed message bits;
Step(iv): For each selected pixel $C_{i,j}$, if its LSB is equal to the message bit, do not change it, otherwise, turn to the next step;

Step(v): Judge whether the value of $C_{i,j}$ is 0 or 255; if so, replace the LSB of $C_{i,j}$ with the message bit, otherwise, turn to the next step;

Step(vi): Constitute the pixel group $y$ with the right horizontal adjacent $C_{i,j+1}$ or the left horizontal adjacent pixel $C_{i,j-1}$, and then determine the category $k_{optimal}$ of the pixel group according to (14);

Step(vii): Obtain $b_{LSB_{optimal}}^{ij}$, then produce a pseudo-random number $n_{ij}$ within the range $(0,1)$, if $n_{ij} \in (0,2b_{LSB_{optimal}}^{ij})$, then $S_{ij} = C_{ij} + 1$; otherwise, $S_{ij} = C_{ij} - 1$. The reason why selecting the threshold $2b_{LSB_{optimal}}^{ij}$ is that $b_{LSB_{optimal}}^{ij} + b_{LSB_{optimal}}^{ij} = 0.5$.

The extraction embedding process of the optimized LSB matching steganography is summarized as follows:

Input: stego-image $\{S_{ij}\}$ which has $N = N_1 \times N_2$ pixels, the rate $\lambda$ of pixels used for embedding, and stego-key $k_1$.

Output: secret message.

Step(i): According to $\lambda$ and the stego-key $k_1$, select the set of image pixels $\{C_{ij}\}$ to embed message;

Step(ii): Select the LSB in each selected pixel $C_{i,j}$, and combine them into the secret message.

In summary, for a cover image, we should calculate $K$ optimized embedding matrices corresponding to each sub-Gaussian distribution. Then the message bits are embedded by adding or subtracting 1 according to the probability indicated by the embedding matrix corresponding to the category the pixel belonging to.

V. EXPERIMENT AND ANALYSIS

In this section, we present some experimental results to demonstrate the effectiveness of our proposed optimized LSB matching steganography compared with the existing LSB matching methods. A common way of testing steganographic schemes is to report the detection metric of some steganalysis methods empirically estimated from a database of cover and stego images where each stego image carries a fixed relative payload[16].

Here the cover image database consists of 1000 images which were downloaded from USDA NRCS Photo Gallery [23]. The images are of very high resolution TIF files (mostly $2100 \times 1500$) and appear to be scanned from a variety of film sources. For testing, the images were resampled to $614 \times 418$ and converted to grayscale (The tool used is Advanced Batch Converter 3.8.20, and the selected interpolation filter is bilinear).

Firstly, the cover images were used to generate 3 groups of 1000 stego images of the existing LSB matching with the relative payload $\lambda = 1$, $\lambda = 0.75$ and $\lambda = 1$, respectively. Then, another 3 groups of 1000 stego images of the optimized LSB matching were generated with the relative payload $\lambda = 1$, $\lambda = 0.75$ and $\lambda = 1$, respectively. The 2000 stego images of two kind of steganography with the same relative payload and the corresponding cover images were used to build a training set and a test set. The training and the test sets were built randomly, both containing 50% cover and 50% stego images. The training sets are used to train some steganalysis detector. The security of two kind steganography with the same relative payload is compared by the detection performance on test sets.

The steganalysis detector makes two types of errors - either detect the cover image as stego (false alarm, or false positive) or recognize the stego image as cover (missed detection, or false negative). The corresponding probabilities are denoted $P_{FA}$ and $P_{MD}$. The receiver operating characteristic (ROC) curve is obtained by plotting $1-P_{MD}(P_{FA})$ as a function of $P_{FA}$. The ROC curve can be reduced into a scalar detection measure called the minimum error probability:

$$P_e = \min_{P_{FA}} \frac{1}{P_{FA}} \left( P_{FA} + P_{MD}(P_{FA}) \right)$$ (15)

Both the ROC curve and the minimum error probability $P_e$ can be used as the detection performance metrics of the steganalysis method for some steganography with the fixed relative payload.

A. ROC curves metric

In the experiments, the first steganalytic detector [2] adopts the Difference Characteristic Function Moments as the features, and Fisher linear discriminator (FLD) as the classifier. The detection performance for the existing LSB matching is excellent [2].

We adopt ROC curves to evaluate the security performances of two steganography methods. The ROC curves of detector show how the detection probability (true positive rate, the fraction of the stego images that are correctly classified) and the false alarm probability (false positive rate, the fraction of the cover images that are misclassified as stego images) vary as detection threshold is varied. The lower the curve is, the more difficult the detector is. It means that the corresponding steganography is more secure.

Figure 2 shows the ROC curves of the existing LSB matching and the optimized LSB matching with the relative payload $\lambda = 1$. Figure 3 shows the ROC curves with $\lambda = 0.75$. Figure 4 shows the ROC curves with $\lambda = 0.5$. From the figures, the ROC curves of the optimized LSB matching are higher than those of the existing LSB matching. Thus the optimized LSB matching is more secure than the existing LSB matching. Further more, the distinctions between the ROC curves of two steganography become small with the decrease of the relative payload. It means that the improvement of security performance is more obvious when $\lambda$ is larger.

B. Minimum error probability metric

We also use the steganalysis detector in the literature[1] to compare the security of two steganography methods.
The steganalysis detector is famous in the realm of steganalysis and steganography. We use the second-order SPAM features with T=3 and first-order SPAM features with T=4 respectively. There are 686 dimensions in the second-order features, and 162 dimensions in the first-order features. With regards to machine learning, we use soft-margin SVMs with a Gaussian kernel of width $\gamma$. Soft-margin SVMs penalize the error on the training set through a hyper-parameter $C$. In this section, the minimum error probability in (15) is used to evaluate the security. The steganography, with larger $P_E$, is more secure than that with smaller $P_E$.

In the experiment, the parameters $C = 300$ and $\gamma = 5$. Note that we didn’t using a grid-search with five-fold cross-validation on the training set to obtain the best $C$ and $\gamma$ as in [1]. The reason is that we only need to find out which steganography has larger $P_E$. We needn’t the best detector with the best $C$ and $\gamma$. The grid-search is time consuming. As a result, the values of $P_E$ in our experiments is larger than that in the literature [1]. But it is trivial to compare the security of two steganographic methods.

Table I demonstrates the security performances of the existing LSB matching steganography and the optimized LSB matching steganography. From table I, with each relative payload, the minimum error probability $P_E$ of the optimized LSB matching is larger than that of the existing LSB matching. So the optimized LSB matching is more secure than the existing one. On the other hand, the distinction of the minimum errors of two steganography decreases when the relative payload become small. Thus the larger the $\lambda$ is, the more the security improvement of the optimized LSB matching is. The results are consistent with the results in V-A section.

<table>
<thead>
<tr>
<th>Relative Payload</th>
<th>SPAM Features</th>
<th>existing LSB Matching</th>
<th>Our optimized LSB matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>Second-order with T=3</td>
<td>0.3955</td>
<td>0.4098</td>
</tr>
<tr>
<td></td>
<td>First-order with T=4</td>
<td>0.2962</td>
<td>0.3105</td>
</tr>
<tr>
<td>0.75</td>
<td>Second-order with T=3</td>
<td>0.3636</td>
<td>0.3972</td>
</tr>
<tr>
<td></td>
<td>First-order with T=4</td>
<td>0.2556</td>
<td>0.2893</td>
</tr>
<tr>
<td>1</td>
<td>Second-order with T=3</td>
<td>0.3359</td>
<td>0.3808</td>
</tr>
<tr>
<td></td>
<td>First-order with T=4</td>
<td>0.2263</td>
<td>0.2643</td>
</tr>
</tbody>
</table>

VI. SUMMARY

We present an optimized LSB matching algorithm. The probabilities of adding or subtracting one for embedding, which is also named as the embedding transferring probabilities, are determined by solving an optimization problem. We demonstrate that Fisher information is the quadratic function of the embedding transferring probabilities. Assuming that the groups of pixels in the cover image can be modeled as GMM, we obtain Gaussian mixture distribution of pixel group for each cover image. Based on it, the optimized embedding transferring probabilities are solved by quadratic programming for each sub-Gaussian distribution. The experiments show that the security performance of this new algorithm is better than the existing LSB matching.

Furthermore, the principle of improving the security in our optimized LSB matching algorithm is different with that of steganography using syndrome coding.
optimized embedding method here can be combined with them. The reason is that many steganographic algorithms based on syndrome coding [12][13] don’t define the specific modification operation (adding or subtracting 1) on cover elements, and our optimized LSB matching algorithm can be utilized to decide which operation to be selected. The combination of two kinds of algorithms may further improve the security.

REFERENCES


Yi-feng Sun was born in Xinxiang, China in 1976. He received the B.S., M.S. and PH.D. degrees in computer science from Zhengzhou information science and technology institute, Zhengzhou, China, in 1999, 2002, 2010. He is presently an instructor with the department of Information Security, Zhengzhou information science and technology institute, Zhengzhou, China. His research interests are in image steganography, image steganalysis, computer vision.

Dan-mei Niu was born in Luoyang, China in 1979. She received the B.S. degree from Henan university of science and technology, Luoyang, China in 2006. She is presently a professor with the department of Information Security, Henan university of science and technology, Luoyang, China. Her research interests are in information security, digital rights management.

Guang-ming Tang was born in Wuhan, China in 1963. She received the B.S., M.S. and PH.D. degrees in information security from Zhengzhou information science and technology institute, Zhengzhou, China, in 1983, 1990, and 2008, respectively. She is presently a professor with the department of Information Security, Zhengzhou information science and technology institute, Zhengzhou, China. Her fields of professional interest are information hiding, watermarking and software reliability. She has published 51 research articles and 3 books in these areas. Ms. Tang is a recipient of Provincial Prize for Progress in Science and Technology.

Zhan-zhan Gao was born in Shijiazhuang, China in 1988. He received the B.S. degree in electronic science and technology from Zhengzhou information science and technology institute, Zhengzhou, China, in 2011. He is pursuing the M.S degree in information security at Zhengzhou information science and technology institute. His research interests include information hiding and information security.
A Novel Robust Zero-Watermarking Scheme Based on Discrete Wavelet Transform

Yu Yang1,2 Min Lei1,2
1. Information Security Center, Beijing University of Posts and Telecommunications, Beijing, China
2. National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing, China
Email: yangelm@vip.sina.com, byleimin@tom.com

Huaqun Liu
School of Information and Mechanical Engineering, Beijing Institute of Graphic Communication, Beijing, China

Yajian Zhou1,2 Qun Luo1
1. Information Security Center, Beijing University of Posts and Telecommunications, Beijing, China
2. National Engineering Laboratory for Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing, China

Abstract—In traditional watermarking algorithms, the insertion of watermark into the original signal inevitably introduces some perceptible quality degradation. Another problem is the inherent conflict between imperceptibility and robustness. Zero-watermarking technique can solve these problems successfully. But most existing zero-watermarking algorithm for audio and image cannot resist against some signal processing manipulations or malicious attacks. In the paper, a novel audio zero-watermarking scheme based on discrete wavelet transform (DWT) is proposed, which is more efficient and robust. The experiments show that the algorithm is robust against the common audio signal processing operations such as MP3 compression, re-quantization, re-sampling, low-pass filtering, cutting-replacement, additive white Gaussian noise and so on. These results demonstrate that the proposed watermarking method can be a suitable candidate for audio copyright protection.

Index Terms—zero-watermarking, discrete wavelets transform, copyright protection

I. INTRODUCTION

Recently, the rapid development of the Internet has increased multimedia services, such as electronic commerce, pay-per-view, video-on-demand, electronic newspapers, and peer-to-peer media sharing. As a result, multimedia data can be obtained quickly over high speed network connections. However, the authors, publishers, owners and providers of multimedia data are reluctant to grant the distribution of their documents in a networked environment because the ease of intercepting, copying and redistributing electrical data in their exact original form encourages copyright violation. Therefore, it is crucial for the future development of networked multimedia systems that robust methods are developed to protect the intellectual property right of data owners against unauthorized copying and redistribution of the material made available on the network. Classical encryption systems do not completely solve the problem of unauthorized copying because once encryption is removed from a document, there is no more control of its dissemination.

Digital watermarking is a good approach for providing copyright protection of digital contents (for example text, word, image, audio, video) [1-14]. This technique is based on direct embedding the watermarking information into the digital contents. The embedding operations should not be introduced any perceptible distortions ideally, there must be no perceptible difference between the watermarked and the original version. That is to say the watermark data should be embedded imperceptibly into the audio media. Apart from imperceptibility, capacity and robustness are two fundamental properties of audio watermarking schemes. To make tradeoff between imperceptible and robust, it’s a common method to incorporate the human auditory system (HAS) into a watermarking system. The watermark should be extractable after various intentional and unintentional attacks. These attacks may include additive noise, re-sampling, MP3 compression, low-pass filtering, re-quantization, and any other attack which removes the watermark or confuse the watermark extraction system. Considering a trade-off between capacity, transparency and robustness is the main challenge for audio watermarking applications.

We focus on the audio watermarking scheme in this paper. In traditional audio watermarking techniques, either in spatial domain, transform domain, or dual
domain [15-16], the embedding of watermark into the original audio inevitably introduces some audible quality degradation. Another problem is the inherent conflict between the imperceptibility and robustness. Then, zero-watermarking technique was proposed by some researchers to solve these problems [17–25]. Instead of embedding watermark into the original signal, the zero-watermarking approach just constructs a binary pattern based on the essential characteristics of the original signal and uses them for watermark recovery. In [21], the property of the natural images that the vector quantization indices among neighboring blocks tend to be very similar was utilized to generate the binary pattern. A scheme that combined the zero-watermarking with the spatial-domain-based neural networks was proposed in [22], in which the differences between the intensity values of the selected pixels and the corresponding output values of the neural network model were calculated to generate the binary pattern. Some low-frequency wavelet coefficients were randomly selected from the original image by chaotic modulation and used for character extraction in [23]. And in [24], two zero-watermarks were constructed from the original image. One was robust to signal process and central cropping, which was constructed from low-frequency coefficients in discrete wavelet transform domain and the other was robust to general geometric distortions as well as signal process, which was constructed from DWT coefficients of log-polar mapping of the host image. There are novel zero-watermarking algorithms for audio now. An efficient and robust zero-watermarking technique for audio signal is presented in [25]. The multi-resolution characteristic of discrete wavelet transform (DWT), the energy compression characteristic of discrete cosine transform (DCT), and the Gaussian noise suppression property of higher-order cumulant are combined to extract essential features from the original audio signal and they are then used for watermark recovery. Simulation results demonstrate the effectiveness of the scheme in terms of inaudibility, detection reliability, and robustness. However, all these zero-watermarking techniques are designed for still image and their robustness against some signal processing manipulations or malicious attacks is not satisfying. In this paper, a novel robust zero-watermarking technique for audio signal is proposed.

The paper is organized as follows. Section II introduces basic concepts for discrete wavelet transform. Section III describes a novel robust zero-watermarking scheme. Section IV exhibits the experimental results illustrating that the proposed method can get good robustness. A brief conclusion can be available in Section V.

II. DISCRETE WAVELET TRANSFORM

Wavelet transform is a time domain localized analysis method with the window’s size fixed and convertible. It supplies high temporal resolution in high frequency part, and high frequency resolution in low frequency part of signals. It can distill the information from signal effectively.

For audio, we need convert the 1-dimensional audio into 2-dimensional. Wavelet transform decomposes an audio into a set of band limited components which can be reassembled to reconstruct the original audio without error. Since the bandwidth of the resulting coefficient sets is smaller than that of the original audio, the coefficient sets can be down sampled without loss of information. Reconstruction of the original signal is accomplished by up sampling, filtering and summing the individual sub-bands. For 1-D audio, we convert the 1-D audio into 2-D signal, and then apply DWT corresponds to processing the audio by 2-D filters in each dimension. The filters divide the input audio into four non-overlapping multi-resolution coefficient sets, a lower resolution approximation audio (LL1) as well as horizontal (HL1), vertical (LH1) and diagonal (HH1) detail components. The sub-band LL1 represents the coarse-scale DWT coefficients while the coefficient sets LH1, HL1 and HH1 represent the fine-scale of DWT coefficients. To obtain the next coarser scale of wavelet coefficients, the sub-band LL1 is further processed until some final scale N is reached. When N is reached we will have 3N+1 coefficient sets consisting of the multi-resolution coefficient sets LLN and LHx, HLx and HHx where x ranges from 1 until N. Figure 1 show the representation of 2-levels of wavelet transform.

III. AUDIO WATERMARKING ALGORITHM

A. Embedding Process

Let $A$ be the original audio signal and let $W = \{w(i, j)\}_{i=1,2,..,M_1, j=1,2,..,M_2}$ be the binary-valued image watermark to be embedded. Now we describe the watermark embedding procedure. (Figure 2)
Step 1: Dimension reduction of watermark image. Because the audio signal is one-dimensional and watermarking $W$ is two-dimensional. Therefore the watermarking should be reduced into one-dimensional signal:

$$w = \{w(i) = w(m_i, m_j) | 0 \leq m_i \leq M_1, 0 \leq m_j \leq M_2, i = M_1 \times M_2\}$$

Step 2: Let the original audio $A = (a_1, a_2, ..., a_N)$ with $N$ PCM (pulse-code modulation) samples be segmented into $M = \lceil N / 512 \rceil$ blocks $B_i$, where $M = M_1 \times M_2, i = 1, 2, ..., M$. Each block includes 512 samples.

Step 3: for $i = 1: M$

3.1). The 3-level discrete wavelet packet decomposition is applied to the audio block $B_i$, and can get the approximate sub-band coefficients $C_i = \{c_i(1), c_i(2), ..., c_i(64)\}$, which has 64 samples.

3.2). Computes

$$t(i) = \sum_{j=1}^{64} c_i(j) \times c_i(j) \times j,$$

$$t'(i) = \sum_{j=1}^{64} c_i(j) \times c_i(j)$$

According to the approximate sub-band coefficients $C_i = \{c_i(1), c_i(2), ..., c_i(64)\}$. And gets:

$$s'(i) = \lceil t(i) / t'(i) \rceil.$$  

3.3). Computes $s(i) = s'(i) \mod 2$.

3.4). End

Step 4: We can get the watermarking extraction secret key $k(i) = w(i) \oplus s(i)$ through computing the exclusive or operation between $s(i)$ and watermarking information $w(i)$.

B. Extraction Process

The watermark recovery procedure can be carried out as follows (Figure 3).

Step 1: Let the watermarked audio $A' = (a'_1, a'_2, ..., a'_N)$ be segmented into $M = \lceil N / 512 \rceil$ blocks $B'_i$, where $M = M_1 \times M_2, i = 1, 2, ..., M$. Each block includes 512 samples.

Step 2: for $i = 1: M$

2.1). The 3-level discrete wavelet packet decomposition is applied to the audio block $B'_i$, and can get the approximate sub-band coefficients $C'_i = \{c'_i(1), c'_i(2), ..., c'_i(64)\}$ which has 64 samples.

2.2). Computes
according to the approximate sub-band coefficients \( C'_j = \{c'_1(1), c'_1(2), \ldots, c'_1(64)\} \). And gets

\[
S'(i) = \left[ \frac{T(i)}{T'(i)} \right].
\]

2.3) Computes \( S(i) = S'(i) \text{mod}\ 2 \).

2.4) End

Step 3: We can get the watermarking information \( w'(i) = k(i) \oplus S(i) \) through computing the exclusive or operation between \( S(i) \) and he watermarking extraction secret key \( k(i) \).

Step 4: We can get the binary-valued image watermark according to the extracted watermarking information.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this experiment, two binary stamp image with size \( 32 \times 32 \) (i.e., \( M_1 = M_2 = 32 \)), displayed in Fig. 5, are taken as the original watermark. And we evaluate the performance of our proposed watermarking method for different types of 16 bit mono audio signals sampled at 44.1 KHz as shown in Fig. 4. The sound files are: (a) Pop, (b) Speech, (c) Classic. Each audio file contains 524,288 samples with duration of 11 seconds.

![Figure 4: The signal of original audio](image)

![Figure 5: Watermarks with size 32×32](image)

In order to evaluate the quality of watermarked audio, the following signal-to-noise ratio (SNR) equation is used:

\[
SNR = 10 \log_{10} \frac{\sum_{n=1}^{N} Y^2(n)}{\sum_{n=1}^{N} [Y(n) - Y'(n)]^2}
\]

where \( Y(n) \) and \( Y'(n) \) are original audio signal and watermarked audio signal respectively.

In order to evaluate the robustness of the watermarked algorithm, the following normalized coefficient (NC) and bit error rate (BER) are employed respectively:

\[
NC(W, W') = \frac{\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} W(i, j) W'(i, j)}{\sqrt{\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} W(i, j)^2} \times \sqrt{\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} W'(i, j)^2}}
\]

\[
BER = \frac{\left( \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} |W(i, j) - W'(i, j)| \right) / (M_1 \times M_2)}
\]

where \( B \) is the number of erroneously extracted bits and \( M_1 \times M_2 \) is the size of watermark.

And, the bit error rate (BER) was employed to measure the robustness of our algorithm,

\[
BER = \frac{B}{M_1 \times M_2} \times 100\%
\]

A. Imperceptibility Analysis

One of the main requirements of audio watermarking techniques is inaudibility of the embedded watermark. For the proposed scheme, this requirement is naturally achieved because the watermark is embedded into the secret key but not the original audio signal itself. Actually, the watermarked audio is the identical to the original one.

B. Robustness Test and Analysis

In order to test the robustness of our proposed method, eight different types of attacks is as following:

1. Noise addition: 20 dB additive white Gaussian noise (AWGN) is added to the watermarked audio signal
2. Re-sampling: The watermarked signal originally sampled at 44.1kHz is re-sampled at 22.050kHz, and then restored by sampling again at 44.1 kHz.
3. Low-pass filtering: cut-off frequency is 11.025kHz.
4. Re-quantization: the 16 bit watermarked audio signal is quantized down to 8 bits/sample and again re-quantized back to 16 bits/sample.
5. MP3 compression: MPEG-1 layer 3 compression with 64 kbps is applied to the watermarked signal.
6. MP3 compression: MPEG-1 layer 3 compression with 32 kbps is applied to the watermarked signal.
7. MP3 compression: MPEG-1 layer 3 compression with 128 kbps is applied to the watermarked signal.
8. Noise reduction: Preset shape is Hiss removal.

Table 1 show the NC and BER of the proposed watermarking method in terms of robustness against several kinds of attacks applied to three different types of audio signals.
watermarked audio signal “Speech”, “Classic” and “Pop” respectively.

<table>
<thead>
<tr>
<th>Audio Attack</th>
<th>Speech</th>
<th>Classic</th>
<th>Pop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NC</td>
<td>BER</td>
<td>Extracted image</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
<tr>
<td>8</td>
<td>0.9991</td>
<td>0.001</td>
<td>E</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0</td>
<td>E</td>
</tr>
</tbody>
</table>

Table 2 Comparison between our algorithm and other algorithm

<table>
<thead>
<tr>
<th>Attack</th>
<th>NC(our)</th>
<th>NC([25])</th>
<th>NC([26])</th>
<th>NC([27])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equalization</td>
<td>1</td>
<td>0.97</td>
<td>-</td>
<td>0.998</td>
</tr>
<tr>
<td>MP3 compression(32kbps)</td>
<td>1</td>
<td>0.99</td>
<td>-</td>
<td>0.999</td>
</tr>
<tr>
<td>Low-pass filtering</td>
<td>1</td>
<td>1</td>
<td>0.9971</td>
<td>0.824</td>
</tr>
<tr>
<td>Noise addition</td>
<td>1</td>
<td>1</td>
<td>0.9668</td>
<td>0.999</td>
</tr>
<tr>
<td>Re-sampling</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>0.981</td>
</tr>
</tbody>
</table>

The Table 2 compares our algorithm and other algorithms, we can easily get the robustness of our algorithm is more stronger than other algorithms.

V CONCLUSION

A novel audio zero-watermarking scheme based on discrete wavelet transform (DWT) is proposed in this paper, which is more efficient and robust. The experiments show that the proposed algorithm has good robustness against the common audio signal processing operations such as MP3 compression, re-quantization, re-sampling, low-pass filter, cutting-replacement and additive white Gaussian noise. These results demonstrate that our algorithm can be a suitable candidate for audio copyright protection.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant NO. 60973146, NO. 61170272, NO. 61170269, NO. 61003285, and Institute Level Project Funded by Beijing Institute of Graphic Communication under Grant NO. E-6-2012-23.
REFERENCES


Abstract—There are kinds of multimedia can be accessed conveniently in social networks. Some of multimedia may be used to hiding harmful information. We consider the problem of estimating the stego key used for hiding using least significant bit (LSB) paradigm, which has been proved much difficult than detecting the hidden message. Previous framework for stego key search was provided by the theory of hypothesis testing. However, the test threshold is hard to determine. In this paper, we propose a new method, in which the correct key can be identified by inspecting the difference between stego sequence and its shifted sequence on the embedding path. It’s shown that the new technique is much simpler and quicker than that previously known.

Index Terms—social networks, multimedia contents security, steganography, stego key estimation

I. INTRODUCTION

The aim of steganography is to hide information imperceptibly into cover objects, such as digital images. To hide a message, some features of the original image, also called the cover image, which are chosen by a stego key are slightly modified by the embedding technique to obtain the stego image. Before embedding, the message is usually encrypted\(^3\). Steganography is significant to information security. However, steganographic technique can be used unlawfully by criminals and terrorists, especially in social networks.

Steganalysis aims to break steganography. steganalysis can be classified into two categories: active and passive\(^\[5\]\). The goal of the active steganalysis is to estimate some parameters (stego key, message length etc.) of the embedding algorithm or the hidden message\(^\[3\][4]\), while passive steganalysis deals with identifying the presence/absence of a hidden message or the embedding algorithm used\(^\[5\][6]\). In this paper, we investigate active steganalysis since the aim is to estimate the stego key under the assumptions that we already know the steganographic algorithm.

Trivedi et al\(^\[7\][8]\) presented a method for secret key estimation in sequential steganography. The authors used a sequential probability ratio test to determine the embedding key, which was, in their interpretation, the beginning and the ending of the subsequence modulated during embedding. In fact, sequential embedding is typically used for watermarking. Fridrich et al\(^\[9\][10]\) considered a more typical situation for a steganographic application, in which the key determined a pseudo-randomly ordered subset of all indices in the cover image to be used for embedding. They performed chi-square test to estimate the correct key. Unfortunately, chi-square test suffered from the drawback of no simple way to choose the test threshold to attain a desired target performance.

In this paper, we propose a novel approach to search the stego key used in LSB steganography. The key can be determined by simple count. In Section II, we include a description of the LSB embedding algorithm. In Section III, we give the definitions used in our method, and then describe the method of identifying the correct key. In Section IV, we give two kinds of experiments. we verify our method in the first experiment and show the effectiveness of our new method by attacking steganographic software in the second experiment. Finally, we outline some further directions for research.

II. LSB STEGANOGRAPHY

For an image \(I\) sized \(M \times N\), let the pixel value at \((i, j)\) be \(I(i, j), i \in \{1, 2, \ldots, M\}, j \in \{1, 2, \ldots, N\}\). \(I(i, j) \in \{0, 1, \ldots, 255\}\). Let \(K\) be the space of all possible stego keys. For each key \(K \in K\), let \(Path(K)\) denote the ordered set of element indices visited along the path generated from the key \(K\). When embedding \(\ell\) message bits, the elements in the sequence \(\{I(i, j), (i, j) \in Path(K)\}\) are chosen orderly to be modified by the embedding operation, which is shown in Table I. After embedding, the stego image \(\hat{I}\) with embedding ratio \(q = \ell/MN\) is obtained.

Our task is to identify the stego key \(K\) under the condition that we have a complete knowledge of the embedding algorithm and one stego image.
A. Definitions
To explain the detail of our new technique, we’ll first briefly explore the definitions used in our method.

A non-boundary pixel \( I(i, j) \) has eight adjacent pixels in image: \( I(i-1,j), I(i-1,j), \ldots, I(i+1,j+1) \). The differences between center pixel and its adjacent pixels are denoted by:

\[
\begin{align*}
D_1(i, j) &= I(i-1, j-1) - I(i, j), \\
D_2(i, j) &= I(i-1, j) - I(i, j), \\
& \vdots \\
D_8(i, j) &= I(i+1, j+1) - I(i, j)
\end{align*}
\]

According to \( D_k, k = 1, 2, \ldots, 8 \), we define three kinds of pixel sets—\( H \), \( L \), and \( M \):

- High pixel set: \( I(i, j) \in H \Leftrightarrow \max D_k < 0 \)
- Low pixel set: \( I(i, j) \in L \Leftrightarrow \min D_k > 0 \)
- Middle pixel set: \( I(i, j) \in M \Leftrightarrow \min D_k \leq 0 \) \& \( \max D_k \geq 0 \)

According to the parity of center pixel value, three sets can be divided into six pixel subsets further: \( H^e \), \( H^o \), \( L^e \), \( L^o \), \( M^e \) and \( M^o \). When the center pixel is modified by LSB embedding, there are ten movements among the six subsets, which are shown in Fig.1.

![Figure 1. Movements of each kind of pixels after LSB embedding.](image)

But only four movements, which span different pixel sets, will change the pixel frequencies of set \( H \), \( L \), and \( M \). The four very movements are:

\[
\begin{align*}
H^o (\max D_k = -1) & \square M^e (\max D_k = 0), \\
L^e (\min D_k = 1) & \square M^o (\min D_k = 0)
\end{align*}
\]

where \( X(y) \) denotes pixels belonging to set \( X \) (called \( X \) pixels) that meets requirement \( y \). To simplify the notation, we use simple symbols to denote the corresponding pixel set, shown in TableI.

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Pixel subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>( HM )</td>
<td>( H^e (\max D_k = -1) )</td>
</tr>
<tr>
<td>( LM )</td>
<td>( L^o (\min D_k = 1) )</td>
</tr>
<tr>
<td>( MH )</td>
<td>( M^e (\max D_k = 0) )</td>
</tr>
<tr>
<td>( ML )</td>
<td>( M^o (\min D_k = 0) )</td>
</tr>
</tbody>
</table>

Finally, we define a flipping operation on sequence \( \{ \hat{I}(i, j), (i, j) \in Path(K), \hat{I}(i, j) \in \hat{I} \} \):

Flipping Operation: \( 0 \leftrightarrow 1, 2 \leftrightarrow 3, \ldots, 254 \leftrightarrow 255 \)

Then a shifted sequence \( \{ \hat{I}(i, j), (i, j) \in Path(K) \} \) is obtained.

B. Method for Identifying Correct Key
Let \( \hat{F}_H(K) \) and \( \hat{F}_L(K) \) be the frequency of \( X \) pixels in \( \{ \hat{I}(i, j), (i, j) \in Path(K) \} \) and \( \{ \hat{I}(i, j), (i, j) \in Path(K) \} \) respectively.

Obviously:

\[
\begin{align*}
&\hat{F}_M(K) = \hat{F}_H(K) - \hat{F}_L(K) \\
&\hat{F}_M(K) = \hat{F}_H(K) - \hat{F}_L(K)
\end{align*}
\]

The frequency change of \( H \) is independent of that of \( L \). Therefore, we only consider the frequency of \( M \) pixels. It’s apparent that:

\[
\begin{align*}
&\hat{F}_{HM}(K) = \hat{F}_H(K) - \hat{F}_{MH}(K) - \hat{F}_{ML}(K) \\
&\hat{F}_{LM}(K) + \hat{F}_{ML}(K)
\end{align*}
\]

Let \( \alpha \) and \( \beta \) be the frequency of “1” and “0” in message bits and \( \lambda \) be the embedding ratio on the path generated from \( K \), which can be defined as \( \lambda = n/|\text{path}(K)| \), where \( n \) denotes number of embedded message bits along \( \text{path}(K) \), and \( |\text{path}(K)| \) denotes number of pixels in \( \text{path}(K) \). Let \( S(K) \) denote the frequency difference of \( M \) pixel between \( \{ \hat{I}(i, j), (i, j) \in Path(K) \} \) and \( \{ \hat{I}(i, j), (i, j) \in Path(K) \} \), namely: \( S(K) = \hat{F}_M(K) - \hat{F}_M(K) \). So:

\[
\begin{align*}
S(K) &= \hat{F}_M(K) - \hat{F}_M(K) \\
&= \hat{F}_{HM}(K) + \hat{F}_{LM}(K) - \hat{F}_{MH}(K) - \hat{F}_{ML}(K) \\
&= \alpha \lambda F_{MH}(K) + (1 - \beta \lambda) F_{LM}(K) + \beta \lambda F_{MH}(K) + (1 - \alpha \lambda) F_{LM}(K) \\
&= \alpha \lambda F_{MH}(K) - (1 - \beta \lambda) F_{MH}(K) - \alpha \lambda F_{MH}(K) - (1 - \beta \lambda) F_{LM}(K)
\end{align*}
\]
Usually, the message is encrypted, so the message bits are i.i.d. realizations of a binary random variable uniformly distributed on \( \{0,1\} \), therefore, \( \alpha = \beta = 1/2^{[11]} \), hence:

\[
S(K) = (F_{\text{MH}}(K) - F_{\text{IM}}(K) + F_{\text{ML}}(K) - F_{\text{LM}}(K))(\lambda - 1) \tag{4}
\]

We use \( \lambda_0 \) and \( \lambda_1 \) denote the embedding ratio along the correct path and incorrect path respectively. Because the number of embedded message bits along the correct path is \( \ell \), and the number of pixels along correct path is \( \ell \) too, the embedding ratio along the path generated from correct key \( K_i \in K \) is \( \lambda_0 = 1 \). We have:

\[
S(K_i) = (F_{\text{MH}}(K) - F_{\text{IM}}(K) + F_{\text{ML}}(K) - F_{\text{LM}}(K))(\lambda_0 - 1) = 0 \tag{5}
\]

ZHAI et al. have proved that \( D_x \) follows Generalized Gaussian distribution with mean “zero” in [11], therefore:

\[
F_{\text{MH}}(K) > F_{\text{IM}}(K), \quad F_{\text{ML}}(K) > F_{\text{LM}}(K) \tag{6}
\]

If the stego image is not fully embedded, the number of embedded message bits along the incorrect path is smaller than \( \ell \), so the embedding ratio along the path generated from incorrect key \( K_j \in K \) is smaller than 1, namely \( \lambda_j < 1 \). Thus:

\[
S(K_j) = (F_{\text{MH}}(K) - F_{\text{IM}}(K) + F_{\text{ML}}(K) - F_{\text{LM}}(K))(\lambda_j - 1) < 0 \tag{7}
\]

From (5) and (7), we get:

\[
S(K_i) > S(K_j) \tag{8}
\]

Therefore, the difference between correct key and incorrect key is obtained. By calculating \( S(K), K \in K \), the key \( K \) with maximal value of \( S(K) \) can be determined as the correct stego key.

IV. EXPERIMENTS AND ANALYSIS

A. Experiments on Typical Images

We perform some experiments to verify (8). We test 100 typical images sized in 512×512, downloaded from http://sipi.usc.edu/services/database/database.html. We simulate the course of LSB embedding using a Matlab routine, in which the key space is \( 2^{20} \). For embedding ratio \( q = 0.60 \), we embed encrypted message into each image, and calculate the value of \( S(K) \). If the stego image is not fully embedded, the number of embedded message bits along the incorrect path is smaller than \( \ell \), so the embedding ratio along the path generated from incorrect key \( K_j \in K \) is smaller than 1, namely \( \lambda_j < 1 \). Thus:

\[
S(K_j) = (F_{\text{MH}}(K) - F_{\text{IM}}(K) + F_{\text{ML}}(K) - F_{\text{LM}}(K))(\lambda_j - 1) < 0 \tag{7}
\]

From (5) and (7), we get:

\[
S(K_i) > S(K_j) \tag{8}
\]

Therefore, the difference between correct key and incorrect key is obtained. By calculating \( S(K), K \in K \), the key \( K \) with maximal value of \( S(K) \) can be determined as the correct stego key.

B. Experiment on Typical LSB Steganographic Software

We conduct this experiment by using “hide and seek 4.1”[12], which is a popular LSB steganographic software. It uses GIF image containing 320×480 pixels as the cover image and adopts random (num) of Borland C++ 3.1 as the generator. The maximal embedding message length is limited to 19000 byte The initialized state of random (num) is a seed with 16 bits, and “num” is used to control maximal migration step, which is related to message and pixel number that haven’t been embedded. So the stego key of Hide and Seek 4.1 consists of the seed and message length. Therefore, we should recover the seed and the exact length of message.

We convert 100 images, downloaded from http://www.cs.washington.edu/research/magedatabase/gro undtruth/, into GIF format with 320×480 pixels and insert undtruth/, into GIF format with 320×480 pixels and insert different length of message using Hide and Seek 4.1. We perform the experiment with our stego key recovery algorithm described as follows.

0. Estimate the embedding ratio using the method presented in [13] and calculate the possible length of secret message \( \ell = [\ell_{\text{min}}, \ell_{\text{max}}] \).

1. For each \( \ell_k \in [\ell_{\text{min}}, \ell_{\text{max}}] \), test each seed \( K_s \in [0, 2^k - 1] \) of the seed generator used in Hide and Seek 4.1 with \( K_{sk} = (K, \ell_k) \), generate the embedding pixel set \( \hat{I}(i,j), (i,j) \in \text{Path}(K_{sk}) \) .

2. Calculate \( S(K_{sk}) \) on sequence \( \hat{I}(i,j), (i,j) \in \text{Path}(K_{sk}) \).

3. Let \( S_{\text{max}} = \max \{S(K_s) \} \) and \( T = \{K, \ell_k \} | K_s \in [0, 2^k - 1] \) \& \( \ell_k \in [\ell_{\text{min}}, \ell_{\text{max}}] \& S(K_{sk}) = S_{\text{max}} \).

4. If \( |T| = 1 \), then the seed and the message length in \( T \) are declared as the correct seed and length. Otherwise, the algorithm couldn’t find any correct key.

Because our algorithm uses the method of [13] to
estimate the message length, where the possible deviation for the estimation of embedding ratio is $[-0.02, 0.02]$. The effective length of stego key we should test actually is $16 + \log_{10} 0.04\%$. Table III shows the results of our method, where $Pr(\text{succ})$ means the probability of success, defined as: $Pr(\text{succ})$=number of images whose stego key is recovered/ total number of images.

**TABLE III.**

<table>
<thead>
<tr>
<th>Embedding ratio $q$</th>
<th>$Pr(\text{succ})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0</td>
</tr>
<tr>
<td>0.005</td>
<td>12.4%</td>
</tr>
<tr>
<td>0.06</td>
<td>89.3%</td>
</tr>
<tr>
<td>0.08</td>
<td>100%</td>
</tr>
<tr>
<td>0.50</td>
<td>100%</td>
</tr>
<tr>
<td>0.91</td>
<td>100%</td>
</tr>
<tr>
<td>0.95</td>
<td>84.4%</td>
</tr>
<tr>
<td>0.97</td>
<td>44.2%</td>
</tr>
<tr>
<td>0.99</td>
<td>0</td>
</tr>
</tbody>
</table>

From Table III, we can see the performance of our method in recovering the stego key of LSB steganographic software. When $0.08 \leq q \leq 0.91$, our method can recover the stego key of all the images. Namely when $0.08 \leq q \leq 0.91$, the probability of $S(K_0) > S(K_j)$ is 1. But when the embedding ratio approaches to 0 or 1, our method is disabled. Because when $q \rightarrow 0$, the number of data is not enough to recovery the stego key, while $q \rightarrow 1$, the embedding ratio difference between correct path and incorrect path is so small that the correct key can’t be detected. In addition, the embedding message length influences our stego key search ratio most, because all the elements along the path should be calculated. The larger the message length, the slower the search rate. The testing speed is about 43-14500 keys per second.

**V. CONCLUSIONS AND FUTURE WORK**

We have proposed a novel and simple way in the stego key recovery of LSB steganography which can produce satisfying performance in images.

In the future work, we should consider how to reduce computational time and improve the success ratio in searching stego key of images with larger or smaller embedding rate. Finally, we should transfer the method into the spatial sequential LSB steganography, whose stego key can be considered as the beginning and the ending of the message during embedding.

**REFERENCES**


Jing LIU was born in Xuancheng, China in 1985. She received the B.S., M.S. degrees in information security from Zhengzhou information science and technology institute, Henan, China, in 2007 and 2010 respectively. Currently, she is pursuing the Ph.D. degree in information security at Zhengzhou information science and technology institute. Her research interest includes information hiding and information security.
Guangming Tang was born in Wuhan, China in 1963. She received the B.S., M.S., and Ph.D. degrees in information security from Zhengzhou information science and technology institute, Henan, China, in 1983, 1990, and 2008, respectively. Her fields of professional interest are information hiding, watermarking and software reliability.

She is presently a professor with the department of Information Security, Zhengzhou information science and technology institute, Henan, China. She has published 51 research articles and 3 books in these areas.

Ms. Tang is a recipient of Provincial Prize for Progress in Science and Technology.
Facial Expression Spacial Charts for Describing Dynamic Diversity of Facial Expressions

H. Madokoro and K. Sato
Department of Machine Intelligence and Systems Engineering, Faculty of Systems Science and Technology, Akita Prefectural University, Yurihonjo, Japan
Email: {madokoro, ksato}@akita-pu.ac.jp

Abstract—This paper presents a new framework to describe individual facial expression spaces, particularly addressing the dynamic diversity of facial expressions that appear as an exclamation or emotion, to create a unique space for each person. We name this framework Facial Expression Spatial Charts (FESCs). The FESCs are created using Self-Organizing Maps (SOMs) and Fuzzy Adaptive Resonance Theory (ART) of unsupervised neural networks. For facial images with emphasized sparse representations using Gabor wavelet filters, SOMs extract topological information in facial expression images and classify them as categories in the fixed space that are decided by the number of units on the mapping layer. Subsequently, Fuzzy ART integrates categories classified by SOMs using adaptive learning functions under fixed granularity that is controlled by the vigilance parameter. The categories integrated by Fuzzy ART are matched to Expression Levels (ELs) for quantifying facial expression intensity based on the arrangement of facial expressions on Russell’s circumplex model. We designate the category that contains neutral facial expression as the basis category. Actually, FESCs can visualize and represent dynamic diversity of facial expressions consisting of ELs extracted from facial expressions. In the experiment, we created an original facial expression dataset consisting of three facial expressions—happiness, anger, and sadness—obtained from 10 subjects during 7–20 weeks at one-week intervals. Results show that the method can adequately display the dynamic diversity of facial expressions between subjects, in addition to temporal changes in each subject. Moreover, we used stress measurement sheets to obtain temporal changes of stress for analyzing psychological effects of the stress that subjects feel. We estimated stress levels of four grades using Support Vector Machines (SVMs). The mean estimation rates for all 10 subjects and for 5 subjects over more than 10 weeks were, respectively, 68.6% and 77.4%.

Index Terms—Facial Expression Spatial Charts, SOMs, Fuzzy ART, SVMs, SRS-18.

I. INTRODUCTION

A face sends information of various types. Humans can recognize intentions and emotions from diverse information that is exhibited through facial expressions. Especially for people with whom we share a close relation, we can feel and understand health conditions or moods directly from facial expressions. For the role of facial expressions in human communication, it is desirable to develop advanced interfaces between humans and machines in the future [1].

In the 1970s, from a study to determine how to express emotions related to facial expressions, Ekman and Friesen defined six facial expressions shown by people feeling six basic emotions (happiness, disgust, surprise, sadness, anger, and fear) that are apparently universal among cultures. They described that these are basic facial expressions because their associated emotions are distinguishable with high accuracy. However, real expressions are blended intermediate facial expressions that often show mixtures of two or three emotions. Human beings often express various facial expressions simultaneously. For example, eyes can express crying but the mouth can express a smile when someone is moved by an extremely kind deed. Moreover, the processes of expressive facial expressions contain individual differences such as differences of face shapes among people.

Regarding this difference, Akamatsu described that human faces present diversity of two types: static diversity and dynamic diversity [2]. Static diversity is individual diversity that is configured by facial componental position, size, location, etc., consisting of the eyes, nose, mouth, and ears. We can identify a person and determine their gender and impressions using static diversity. We are able to move facial muscles to express internal emotions unconsciously and sequentially or express emotions as a message. This is called dynamic diversity. Facial expressions are expressed as a shift from a neutral facial expression to one of changed shapes of parts and overall configurations constructed with the face. For studying facial expression analysis, we must consider and understand not only static diversity but also dynamic diversity.

When dealing with dynamic diversity of facial expressions, it is necessary to describe relations between physical parameters of facial pattern changes with expressions and psychological parameters of recognized emotion. Physical parameters of facial expressions are the described facial deformations created by expression under consistent measurements of facial patterns that differ depending according to the size and shape of each person. Ekman and Friesen proposed a Facial Action Coding System (FACS) as a method to describe facial...
expression changes from movements on a facial surface. Viewed comprehensively, the FACS is the most popular and standard method for objective description of facial expressions. It is often used in behavioral sciences and psychology. The FACS was developed originally as a tool to measure facial expressions consisting of anatomical stand-alone Action Units (AUs). The FACS is useful to realize natural and flexible man-machine interfaces in the fields of human cognition and behavior science studies. However, special training is necessary to describe AUs as an observer. Moreover, movements of facial expressions that can not be described as AUs exist in practice because AUs are classified subjectively by an observer to examine numerous quantities of facial expressions. Parameter description methods aside from those of AUs of FACS are examined because FACS systematizes appearances of images as facial expressions, not always as suitable parameters to describe shape changes of the facial surface shown by expressions.

In contrast, psychological parameters can be acquired from tasks of cognitive judgments related to emotions indicating facial expression images to a subject as visual stimulation. Although physical parameters of facial expression patterns are unique in each person, psychological parameters of emotion are universal among humans. The emotion to be recognized shows various attributes according to the degree of physical changes of facial expression patterns such as open or closed eyes and mouth. For estimating the degree of emotion, it is necessary to relate physical variations of individual patterns of facial expressions and psychological variations according to their levels. Moreover, the feature space to describe this amount of changes is necessary to describe spaces based on the common scale in each person because emotion is universal. The fact that expressive facial expressions differ among people is associated with the fact that shapes of faces differ among people. For example, the range within which expressions on the facial surface change according to an emotion differs among people. Akamatsu described that adaptive learning mechanisms are necessary to modify models according to characteristics of expression in each person as a platform of a classification mechanism of emotion [2].

For organizing and visualizing facial expression spaces, this paper presents a novel framework to describe the dynamic diversity of facial expressions. The framework accommodates dynamic changes of facial expressions as topological changes of facial patterns driven by facial muscles of expression. For that reason, the framework is suitable to describe the richness of facial expressions using Expression Levels (ELs). The target facial expressions are happiness, anger, and sadness from the basic six facial expressions to represent expression levels as a chart with axes of each expression quantitatively and visually. From temporal facial expression images, we use Self-Organizing Maps (SOMs) [3] that contain self-mapping characteristics to extract facial expression categories according to expressions. We also use Adaptive Resonance Theory (ART) that contains stability and plasticity that enable classification to integrate categories adaptively under constant granularity. We infer relations between categories created by Fuzzy ART and ELs based on Russell’s circumplex model. This paper presents Facial Expression Spatial Charts (FESCs) to represent dynamic diversity of facial expressions as a dynamic and spatial chart. For the experiment, we created original facial expression datasets including images obtained during 7–20 weeks at one-week intervals from 10 subjects with three facial expressions: happiness, anger, and sadness. Experimental results show that our method can visualize and quantify facial expressions between subjects and temporal changes for creating FESCs in each subject. We use stress measurement sheets to assess temporal changes of stress in each subject for analyzing psychological stress, which includes the subjects that affect facial expressions. We analyze relations between FESCs and psychological stress values. Moreover, we estimate stress levels from FESCs.

This paper consists of the following. We review related work to clarify the position of this study in Section II. In Section III, we define two terms: ELs that quantitatively represent individual facial expression spaces and FESCs that we propose in this paper. In Section IV, we explain our proposed method to capture facial expression images, preprocessing, classification of facial expression patterns with SOMs, integration of facial expression categories with Fuzzy ART, and creation of FESCs based on ELs. We explain our original developed facial expression datasets in Section V. We show results of FESCs for 10 subjects in Section VI. In Section VII, we estimate stress levels using FESCs as an application of our method. Finally, we present conclusions and feature work in Section VIII.

II. RELATED WORK

Approaches dealing with facial expressions have changed from emphasis of spatial factors as an extension of pattern classification for recognition of the six basic facial expressions. Increasingly, temporal changes are analyzed to solve the problems posed by dynamic diversity of facial expressions [4]. In this section, we review related work, particularly addressing the latter approaches. As a study to address facial expression changes and its timing factors, Bassili [5] classified facial expressions using motion feature points captured by markers applied on a human face. However, the influence of motion components is not clear because their method can not control stimulations for dynamic changes of facial expressions in their visual psychological experiment.

In recent studies specifically examining dynamic diversity of facial expressions, Ōhta et al. [6] proposed a method based on a model of facial structure elements. They described facial expression patterns using parameter values of construction of facial muscles matching with facial video images. They created a variable model of facial structural elements of the eyebrows, eyes, and
mouth. Through comparison with a target facial pattern and pre-defined standard patterns, this model can extract expression levels to facilitate recognition of facial expressions. Using expression levels, the intervals and temporal changes can be extracted to reveal patterns in the motion of the face that are related to facial expressions. Moreover, temporal changes of expressions, duration, and termination processes can be detected from increasing and decreasing expression levels. However, setting of standard facial expression patterns is necessary for calculating expression levels. Therefore, this model can only describe the maximum level of the standard facial expression patterns. Moreover, this model includes the drawback that setting the feature points manually on the first frame of facial images is necessary.

Nishiyama et al. [7] proposed facial scores as a framework to interpret dynamic aspects of detailed facial expressions based on timing structures of movements of facial parts. They used modes that are segmented facial expressions according to expression levels as an index to describe fine differences in temporal factors of facial expression changes. After dividing a static or dynamic status of movements of facial parts traced using Active Appearance Models (AAMs), facial expression images are segmented against indicators of movements from temporal subtraction norms of feature vectors. The modes are extracted as four patterns to repeat integration of intervals based on hierarchical clustering defined by segmented sectional distances. The capability of describing facial expression degrees is low, although the temporal resolution is high for representation of timing structures from modes. Moreover, various methods have been proposed to extract facial expression intensity [8]–[16]. Similarly, the spatial resolution is as high as four levels, although the temporal resolution is also high.

The level of facial expressions differs according to mental state, context, situation, etc. We actively use the difference to take facial expression images for a long term in each subject. We are aiming at describing and quantifying individual facial expression spaces from temporal changes of long-term facial expression datasets. The degrees of expression levels differ among people and their feelings at any particular time. We consider creation of facial expression spaces in constant granularity and adaptively, thereby avoiding quantification of the maximum expression level. Moreover, we seek to obtain facial expression images under consideration of the effects of mental status.

III. AROUSAL LEVELS AND FACIAL EXPRESSION SPATIAL CHARTS

In this chapter, we explain ELs that use quantification of individual facial expression spaces and FESCs as a framework to integrate ELs.

A. Expression Levels

As described in this paper, we introduce Expression Levels (ELs) for quantifying facial expression intensity based on the arrangement of facial expressions on Russell’s circumplex model [17], as portrayed in Fig. 1(a). In this model, all emotions are constellation on a two-dimensional space: the pleasure dimension of pleasure–displeasure and the arousal dimension of arousal–sleepiness. The ELs include both features of the pleasure and arousal dimensions. We extract dynamics of facial parts such as eyes, eyebrows, and the mouth as topological changes of facial expressions. Input images are categorized using extracted features of topological changes. The ELs are obtained as sorted categories according to their differences in intensity from expressions that are regarded as neutral facial expressions.

B. Facial Expression Spatial Charts

Facial Expression Spatial Charts (FESCs) are a new framework to describe facial expression spaces and patterns of ELs constituting each facial expression. Facial expression spaces are spatial configurations of each facial expression that are used to analyze semantic and polar characteristics of various emotions portrayed by facial expressions [2]. They represent a correspondence relation between the physical parameters that present facial changes expressed by facial expressions and the psychological parameters that are recognized as emotions.

Psychological parameters can be extracted from psychological experiments to take cognitive decisions related to emotions. Physical parameters must be described based on a certain standard of types and based on the facial deformity that invariably arises from expressions on different facial patterns that differ in each person, as represented by FACS.

Our target facial expressions are happiness in the first quadrant, anger in the second quadrant, and sadness in the third quadrant of Russell’s circumplex model. Fig. 1 shows the correspondence relation between Russell’s circumplex model and an FESC. The value of each axis on the FESC shows the maximum values of ELs. The FESC is created by the connection among maximum values of ELs.

IV. PROPOSED METHOD

A. Whole architecture

Akamatsu described the adaptive learning mechanisms necessary for modification according to individual char-
task feature changes related to expressions. In contrast, feature-based representation demands high calculation costs for the process of extracting and tracing feature points. Moreover, feature-based representation contains problems of precision and stability in cases of numerous samples being processed automatically. For our method, we use view-based representation after converting images with filters of Gabor wavelets showing similar characteristics to those of a human primary visual cortex. Our processing target is to extract ELs from pattern changes of one facial expression from neutral facial expression. Therefore, we consider that the changed parts are apparent on the feature space after converting Gabor wavelets, without tracking of feature points based on AUs.

The period during which images were obtained was expanded from several weeks to several months. We were unable to constrain external factors completely, e.g., through lighting variations, although we took facial expression images in constant conditions. Therefore, in the first step, brightness values are preprocessed with normalization of the histogram to the target images.

In the next step, features are extracted using Gabor wavelet filtering. In the field of computer vision and image processing, information representation of Gabor wavelets is a popular method for an information-processing model based on human visual characteristics. The information representation of Gabor wavelets that can emphasize an arbitrary feature with inner parameters shows the same characteristic of response selectivity in a receptive field.

At the final step, we applied downsampling for noise reduction and compression of the data size. In this method, we set the initial position of the template to contain facial parts for capturing facial images. We use template-matching methods to trace the region of interest of a face in real time. However, the trace results of the region of interest yield errors caused by body motion. These errors can be removed through the procedure of downsampling. The downsampling window that we set is $10 \times 10$ pixels. The dimension of the target images is compressed from $80 \times 90$ pixels to $8 \times 9$ pixels.

C. Classification of facial patterns with SOMs

For classification according to ELs, 200-frame images are normalized in a constant range. In this method, we
used SOMs, which are unsupervised neural networks with competitive learning in neighborhood regions. Fig. 3(a) depicts a network architecture of a SOM. The network architecture of SOMs typically includes two layers: the input layer and the mapping layer. All units on the mapping layer are connected to all units of the input layer while maintaining weights between both layers. When a set of input data is propagated, a unit whose weights are the most similar to the input data is burst. Weights on the burst unit and its neighbor units are updated to be close to the input data, which is the learning of SOMs. Similarity among input data limits the features of topological saving that are reflected in the distance of the burst unit on the one-dimensional or two-dimensional units. According to the progress of learning, similar feature weights are mapped to neighbor units; other units are mapped to separate units. The SOM training algorithm is the following.

1) Let $w_{i,j}(t)$ be the weight from the input unit $i$ to the Kohonen unit $(n, m)$ at time $t$. The weights are initialized with random numbers.

2) Let $x_i(t)$ be the input data to the input unit $i$ at time $t$. The Euclidean distance $d_j$ between $x_i(t)$ and $w_{i,j}(t)$ is calculated as

$$d_j = \sqrt{\sum_{i=1}^{T} (x_i(t) - w_{i,j}(t))^2}.$$  \hspace{1cm} (1)

3) The win unit $c$ is defined, for which $d_j$ becomes a minimum as

$$c = \arg \min (d_j).$$ \hspace{1cm} (2)

4) Let $N_c(t)$ be the units of the neighborhood of the unit $c$. The weight $w_{i,j}(t)$ inside $N_c(t)$ is updated using the Kohonen training algorithm as $(\alpha(t))$ is the training coefficient, which decreases with time. 

$$w_{i,j}(t + 1) = w_{i,j}(t) + \alpha(t)(x_i(t) - w_{i,j}(t)).$$ \hspace{1cm} (3)

5) Training is finished when the iterations reach the maximum number.

D. Integration of facial patterns with Fuzzy ART

The input data are classified in the fixed number of units of the mapping layer. Therefore, classification results are relative. In contrast, classification under the fixed granularity is required for long-term datasets in each subject. In our method, facial expression categories are integrated with Fuzzy ART to learn weights of SOMs.

The use of ART, which was proposed by Grossberg et al., is a theoretical model of unsupervised and self-organizing neural networks forming a category adaptively in real time while maintaining stability and plasticity. Actually, ART has many variations [19]. We use Fuzzy ART [20], into which analog values can be input. Fig. 3(b) depicts a network architecture of SOMs. The network architecture of Fuzzy ART consists of three fields: Field 0 (F0) of receiving input data, Field 1 (F1) for feature representation, and Field 2 (F2) for category representation. The Fuzzy ART algorithm is the following.

E. Allocation of ELs to FESCs

The quantities of neurons of F1 and F2 are, respectively, $m$ and $n$. Actually, $w_i$ represent the weights between respective F2 neurons $i$ and corresponding F1 neurons. All $w_i$ are initialized as one. Fuzzy ART dynamics are determined using a choice parameter $a (a > 0)$, a learning rate parameter $r (0 \leq r \leq 1)$, and a vigilance parameter $p (0 \leq p \leq 1)$.

For each input $x_i$ to each unit $i$ on the F2, the choice function $T_i$ is defined as

$$T_i = \frac{|x_i \land w_i|}{|a + w_i|}. \hspace{1cm} (4)$$

In addition, $i_c$, which is the maximum value $c$ of $T_i$, is selected for a category as a winner. The category with the smallest index is chosen if more than one unit is maximal. When $T_c$ is selected for a category, the $c$-th neuron on the F2 is set to 1; other neurons are set to zero. Resonance or resetting is judged as the following equation if the selected category matches $x_i$. For the activation value $x_i \land w_c$ propagated from the signal of the $c$-th unit on F2 to F1, if the match function is

$$\frac{|x_i \land w_c|}{|a|} \geq p,$$ \hspace{1cm} (5)

then resonance occurs at $x_i$ and $c$. The chosen category is decided and the weights are updated as

$$w_{i0} = r(x_i \land w_{i0}) + (1 - r)w_c.$$ \hspace{1cm} (6)

Therefore, $c$ is reset if resonance does not occur. The unit to contain the next maximum value $T_i$ is chosen again. Resonance or reset is judged again. Fuzzy ART creates a new unit on F2 if all units are reset.

© 2012 ACADEMY PUBLISHER
number is the same as the number of categories integrated using Fuzzy ART.

The maximum values of ELs are depicted to an FESC. The center of an FESC is EL=0, which represents a neutral facial expression. With increasing ELs, facial expression categories are assigned to the outside of the triangle.

F. Stress Estimation with SVMs

In this study, we specifically examine the effects of psychological stress on facial expressions. We measured psychological stress values using a stress checking sheet while taking facial expression images together. As an application of our study, we estimate the stress levels using FESCs. We used SVMs [18], which have high recognition capability, for mapping input data to a high dimensional space using kernel tricks.

Kernel function \( K \) uses the polynomial kernel, the Radial Basis Function (RBF), and the Sigmoid kernel, etc. For this study, we used RBF defined as

\[
K(x, x_i) = \exp\left(-\frac{||x-x_i||^2}{\lambda}\right),
\]

where \( \lambda \) is the variance of RBF. The property of the Kernel differs in the setting of \( \lambda \). Therefore, we evaluate our method using results to change in a certain range.

V. Datasets

For this study, we created an original dataset dealing with long-term facial expression changes. Additionally, for analyzing psychological effects of temporal changes of ELs, we measured psychological stress levels using a stress sheet after taking facial expression images each time.

A. Facial expression dataset

Human show facial expressions of two types: spontaneous facial expressions and intentional facial expressions. Taking a steady and long-term dataset without regard to a camera and a situation is a challenging task, although spontaneous facial expressions present the advantage of corresponding directly to affection or emotion. Moreover, the cause-and-effect relation of facial expressions from emotions is uncertain. In contrast, intentional facial expressions are used as a communication method to communicate something positively to other person, especially in social communication. We set a target to create an original international facial expression datasets to obtain a long-term facial expression dataset for selected subjects. Moreover, intentional facial expression datasets are suitable to keep the number of subjects as a horizontal dataset.

Open datasets of facial expression images are released from some universities and research institutes to be used generally in many conventional studies for performance comparisons of facial expression recognition or automatic analysis of facial expressions [1]. However, the specifications vary in each dataset, and among datasets. As static facial images, the dataset presented by Ekman and Friesen is a popular dataset comprising collected various facial expressions used for visual stimulation in psychological examinations of facial expression cognition. As dynamic facial images, the Cohn–Kanade dataset and the Ekman–Hager dataset are widely used, especially in experimental applications [21]. In recent years, the MMI Facial Expression Database presented by Pantic et al. [22] has become a widely used open dataset containing both static and dynamic images. These dynamic datasets contain a sufficient number of subjects as a horizontal dataset. However, images are taken only once for each person. No dataset exists in which the same subject has been traced over a long term.

According to the mental status, circumstances, and context of a subject, facial expression patterns differ each week, even for the same subject. We set a long term to obtain facial expression images and thereby create individual models of FESCs and for use in estimating stress levels in each subject. Existing methods created an integrated model for classification and recognition of facial expressions, although expression patterns were affected by individual differences. We create individual models in each subject to extract ELs for creating FESCs and to estimate stress levels using each FESC. For development of individual models, we create long-term facial expression datasets compiled with information gathered during periods as long as 20 weeks.

B. Target facial expressions

In our daily life, we frequently observe mixed facial expressions rather than a single facial expression. We believe that the estimation accuracy can be improved if mixed facial expression datasets are used. Considering the load for subjects, it is a challenging task to collect numerous samples of mixed facial expressions. Moreover, the mixture ratio of facial expressions remains unclear. Therefore, the target of this study is particular facial expressions of three types.

Reducing the load to subjects for taking long-term facial expression images, we selected target facial expressions from the six basic facial expressions described by Ekman. As a cultural factor of expression, Ekman reported that Japanese people express a smile even at times when they feel disgust [23]. We considered that this is a difficult facial expression used by Japanese people. Regarding fear, we received opinions that it is extremely difficult to express poses of fear because it arises only from rare situations that are not usually encountered in daily life. For surprise, many subjects, especially men, feel embarrassed, although they do not feel difficulty in expressing surprise. Moreover, mixed facial expressions with happiness are apparent. Considering these restrictions, opinions from subjects, and the assignment of facial expressions on Russell’s circumplex model, we selected happiness, anger, and sadness as target facial expressions.
C. Acquisition of facial expression images

We took images of three facial expressions with 10 subjects over a long term. The terms of taking images differed among subjects, but images were taken during 7–20 weeks at one-week intervals. Details of subjects are five females (Subjects A, B, C, and D were 19; Subject E was 21) and five males (Subjects F and J were 19; Subjects G, H, and I were 22) university students.

We began to take facial expression images when a subject became accustomed to the experimental environment after some trials. Considering generality and usability, we used a USB camera (Qcam, Logicool; Logitec Corp.). We set the environment to simulate a normal indoor condition (lighting by fluorescent lamps). We took frontal facial images to include the region containing facial components such as the eyebrows, eyes, nose, and mouth. We previously indicated to subjects to restrain the head position such as the eyebrows, eyes, nose, and mouth. We previously indicated to subjects to restrain the head position as much as possible. The images were fit to the constant range including the facial region. We used a method using Haar-like features and Boosting for tracking a face region. We used a method using Haar-like features and Boosting for tracking a face region. We used a method using Haar-like features and Boosting for tracking a face region. We used a method using Haar-like features and Boosting for tracking a face region.

We took images of three facial expressions with 10 sets of 200 frames with the sampling rate of 10 frames per second. One set of data consisted of 200 frames with the sampling rate of 10 frames per second.

As conditions of this experiment, we previously explained to all subjects the following preferred procedures:

- show posed facial expressions to the camera;
- avoid movement of the head as much as possible;
- repeat expressions three times during the acquisition time of 20 s;
- make a maximum facial expression;
- return to the neutral facial expression in each interval; and

- relax.

Subjects were not trained using FACS. We agreed with participants that facial expression images shall not be used except for this research. We did not explain details of this study or datasets used for stress level estimation.

D. Stress measurements

For this study, we used stress measurement sheets known as the Stress Response Scale 18 (SRS-18) by Suzuki et al. [25]. The SRS-18 comprises question sheets that can measure responses related to psychological stress easily in a short time and record many that we meet in our daily life. Specific psychological stress responses are gloom, anxiety, and anger (emotional responses), lethargy and difficulty concentrating (cognitive responses), decreased efficiency of work (behavioral responses), etc. caused by stressors. This sheet can measure stress responses according to three factors: dysphoria or anxiety, displeasure or anger, and lassitude. The SRS-18 has 18 questions that can elicit answers of four types: strongly no, no, yes, and strongly yes. The scores for answers correspond respectively to zero to three points. The range of total points is 0–54 points. A high total score indicates a high level of psychological stress. Moreover, four grades of Level 1 (weak), Level 2 (normal), Level 3 (slightly high), and Level 4 (high) are classified from the points. Subjects complete this sheet before taking facial expression images to reduce the effect from stress checking results.

Depending on the subject, we consider that subjective factors affect estimation accuracy because SRS-18 is based on subjective assessments. In this study, we used a salivary amylase monitor produced by Nipro Corp. to measure stress values as inferred from activated salivary amylase. Kashiwano et al. [26] portrays the usability of stress measurements and evaluation using a salivary amylase monitor. For this measurement, subjects are prevented from eating food for two hours beforehand. During the measurement phase, subjects hold a chip in the mouth to take salivary measurements for about 30 s. The load for subjects in this study is very high. Furthermore, the salivary amylase monitor has sensitive response to temporal stress and presents a wide variation of measurement results. In contrast, SRS-18, which consists of only 18 terms as questions, can be completed in a short time. Moreover, SRS-18 can be used to assess various body responses. We consider that SRS-18 is a useful check sheet used in our experiment.

VI. EXPERIMENTAL RESULTS

In this section, we present results of each step to create an FESC to apply to one subject. Subsequently, we present results of FESCs for application to 10 subjects.

A. Results of an FESC (Subject A)

For corresponding to ELs, we extracted categories classified with SOMs and integrated with Fuzzy ART. The
Expression Levels portrays average FESCs of 10 subjects. As the facial images differ among subjects for the 7–20 weeks. Fig. B. Results of FESCs (10 subjects) stress that is assessed as measured using the SRS-18. In the next section, we will analyze these results with that the changes are attributable to psychological effects.

 Changed in each week in the same subject. We consider that the changes are attributable to psychological effects. The result contains slight variation.

 Expression are represented as changes of ELs. The ELs are expressed four times. Start and terminal timings of expression are represented as changes of ELs. The ELs are changed according to the expressions, although the result contains slight variation.

 Figure 6. Transition of ELs in each facial expression (Subject A at ninth week).

 We evaluated temporal changes of ELs to verify the correspondence of ELs and facial expressions. Fig. 6 presents results of temporal changes of ELs of happiness, anger, and sadness. The horizontal axis depicts the frames that consist of 200 frames in each dataset. The vertical axis depicts ELs. We marked the dashed vertical lines to the start and terminal positions of expression. The subject showed expressions of three or four times during one dataset. In this dataset of Subject A at the ninth week, happiness is expressed three times; anger and sadness are expressed four times. Start and terminal timings of expression are represented as changes of ELs. The ELs are changed according to the expressions, although the result contains slight variation.

 Fig. 7 depicts some examples of FESCs. The FESCs show temporal changes of facial expression patterns that changed in each week in the same subject. We consider that the changes are attributable to psychological effects. In the next section, we will analyze these results with stress that is assessed as measured using the SRS-18.

 B. Results of FESCs (10 subjects)

 We created FESCs for 10 subjects. The terms of taking images differ among subjects for the 7–20 weeks. Fig. 8 portrays average FESCs of 10 subjects. As the facial expression of happiness in all average ELs of all subjects, the maximum value is 9.1 in Subject J. The minimum value is 5.6 in Subject G. As facial expressions of anger, Subjects E and F respectively show the maximum and minimum ELs. As facial expressions of sadness, Subjects C and J respectively show the maximum and the minimum ELs. The triangle of the FESC of Subject J is smaller than those of other subjects.

 As sexual differences, Kring et al. [27] portray that emotional expressions with facial expressions of women are richer than those of men. In our experience, most female subjects were not averse to showing intense facial expressions in front of a camera. We inferred that male subjects were shy or not good at making facial expressions. This trend appears on FESC as the size of triangles.

 VII. Estimation of Stress Levels

 The degree of actual facial expressions is modified by various types of psychological effects, a situation, atmosphere, etc., although spontaneous and intentional facial expressions are triggered by emotional changes and intentional social restrictions, such as when one makes a fake smile. In this study, we have acquired facial expression images continually during a long period in an identical situation. The FESCs show various distributions in each week. Therefore, the ELs that show the degrees of expressions in our method differ each week. In this experiment, we specifically examine the effect between expressions and stress from psychology for estimating stress levels from FESCs.

 We used SVMs [18], which have high recognition capability, for mapping input data to a high dimensional space using kernel tricks. We evaluated estimation rates using Leave-One-Out Cross Validation (LOOCV). The estimation targets are stress evaluation values of four
This paper presents FESCs as a framework to describe individual facial expression spaces based on the consideration of facial expressions created by emotion as an individual space in each person. The ELs are created by categories that are classified by SOMs and integrated with Fuzzy ART. The FESCs are created with the axes of ELs of three facial expressions (happiness, anger, and sadness) based on Russell’s circumplex model. We created an original facial expression dataset of 10 subjects (five male subjects and 5 female subjects) for seven weeks. Using this dataset, our method can express individual facial expression spaces based on the consideration of facial expressions created by emotion as an individual space in each person. We consider that estimation performance will be improved if long-term datasets of more than 10 weeks were obtained to continue to obtain vertical datasets.

Using our method, we achieved efficient estimation of stress levels, although we used SVMs under the condition of disproportionate training data distribution. We evaluated all datasets using LOOCV. The number for datasets for each stress level is various. The number for datasets of Level 2 is the largest: about 50%. This rate reaches 80% when including the number of datasets of Level 1. Moreover, five patterns of datasets were produced, which correspond to four subjects; one set of data consisted of stress levels. Six patterns of five subjects produced only two samples. We used all datasets of these few samples except without exception, although it is difficult to learn and to estimate these data using conventional generalization capabilities. To collect these data evenly is a challenging task because stress distributions vary among individuals. We consider that estimation performance will be improved if long-term datasets of more than 10 weeks were obtained to continue to obtain vertical datasets. Four categories from such few data. Therefore, we selected the datasets of these subjects for more than 10 weeks. The mean estimation rate of Subjects A, B, F, G, H, and I is 77.4%. We consider that the estimation performance will be improved if long-term datasets of more than 10 weeks were obtained to continue to obtain vertical datasets.

Using our method, we achieved efficient estimation of stress levels, although we used SVMs under the condition of disproportionate training data distribution. We evaluated all datasets using LOOCV. The number for datasets for each stress level is various. The number for datasets of Level 2 is the largest: about 50%. This rate reaches 80% when including the number of datasets of Level 1. Moreover, five patterns of datasets were produced, which correspond to four subjects; one set of data consisted of stress levels. Six patterns of five subjects produced only two samples. We used all datasets of these few samples except without exception, although it is difficult to learn and to estimate these data using conventional generalization capabilities. To collect these data evenly is a challenging task because stress distributions vary among individuals. We consider that estimation performance will be improved if long-term datasets of more than 10 weeks were obtained to continue to obtain vertical datasets.

VIII. CONCLUSION

This paper presents FESCs as a framework to describe individual facial expression spaces based on the consideration of facial expressions created by emotion as an individual space in each person. The ELs are created by categories that are classified by SOMs and integrated with Fuzzy ART. The FESCs are created with the axes of ELs of three facial expressions (happiness, anger, and sadness) based on Russell’s circumplex model. We created an original facial expression dataset of 10 subjects (five male subjects and 5 female subjects) for seven weeks. Using this dataset, our method can express individual facial expression spaces using FESCs. Moreover, we used SRS-18 for measuring the stress levels of each subject before taking images. We analyzed the effects of psychological stress using FESCs. The results show that happiness and sadness are affected by stress in most subjects.
Future studies must evaluate intentional and spontaneous facial expressions for discrimination using symmetry properties of the horizontal direction to represent facial expression rhythms created by individual patterns of time changes of ELs. Moreover, we will seek to increase the number of subjects for horizontal studies between subjects and to capture long-term datasets for vertical studies in each subject to analyze and to elucidate relations between facial expressions and stress.

ACKNOWLEDGMENT

The authors would like to thank fellow engineers at SmartDesign Co. Ltd. for providing much appreciated advice and for developing a system for taking facial images for these experiments. We also thank the 10 students at our university who participated as subjects by letting us take facial images over such a long period.

REFERENCES


Hirokazu Madokoro received the ME degree in information engineering from Akita University in 2000 and joined Matsushita Systems Engineering Corporation. He moved to Akita Prefectural Industrial Technology Center and Akita Research Institute of Advanced Technology in 2002 and 2006, respectively. He received the PhD degree from Nara Institute of Science and Technology in 2010. He is currently an assistant professor at the Department of Machine Intelligence and Systems Engineering, Akita Prefectural University. His research interests include in machine learning and robot vision. He is a member of the Robotics Society of Japan, the Japan Society for Welfare Engineering, the Institute of Electronics, Information and Communication Engineers, and the IEEE.
Kazuhito Sato received the ME degree in electrical engineering from Akita University in 1975 and joined Hitachi Engineering Corporation. He moved to Akita Prefectural Industrial Technology Center and Akita Research Institute of Advanced Technology in 1979 and 2005, respectively. He received the PhD degree from Akita University in 1997. He is currently an associate professor at the Department of Machine Intelligence and Systems Engineering, Akita Prefectural University. He is engaged in the development of equipment for noninvasive inspection of electronic parts, various kinds of expert systems, and MRI brain image diagnostic algorithms. His current research interests include biometrics, medical image processing, facial expression analysis, computer vision. He is a member of the Medical Information Society, the Medical Imaging Technology Society, the Japan Society for Welfare Engineering, the Institute of Electronics, Information and Communication Engineers, and the IEEE.
Call for Papers and Special Issue Proposals

Aims and Scope.

Journal of Multimedia (JMM, ISSN 1796-2048) is a scholarly peer-reviewed international scientific journal published bimonthly, focusing on theories, methods, algorithms, and applications in multimedia. It provides a high profile, leading edge forum for academic researchers, industrial professionals, engineers, consultants, managers, educators and policy makers working in the field to contribute and disseminate innovative new work on multimedia.

The Journal of Multimedia covers the breadth of research in multimedia technology and applications. JMM invites original, previously unpublished, research, survey and tutorial papers, plus case studies and short research notes, on both applied and theoretical aspects of multimedia. These areas include, but are not limited to, the following topics:

- Multimedia Signal Processing
- Multimedia Content Understanding
- Multimedia Interface and Interaction
- Multimedia Databases and File Systems
- Multimedia Communication and Networking
- Multimedia Systems and Devices
- Multimedia Applications

JMM EDICS (Editors Information Classification Scheme) can be found at [http://www.academypublisher.com/jmm/jmmedics.html](http://www.academypublisher.com/jmm/jmmedics.html).

Special Issue Guidelines

Special issues feature specifically aimed and targeted topics of interest contributed by authors responding to a particular Call for Papers or by invitation, edited by guest editor(s). We encourage you to submit proposals for creating special issues in areas that are of interest to the Journal. Preference will be given to proposals that cover some unique aspect of the technology and ones that include subjects that are timely and useful to the readers of the Journal. A Special Issue is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

The following information should be included as part of the proposal:

- Proposed title for the Special Issue
- Description of the topic area to be focused upon and justification
- Review process for the selection and rejection of papers.
- Name, contact, position, affiliation, and biography of the Guest Editor(s)
- List of potential reviewers
- Potential authors to the issue
- Tentative time-table for the call for papers and reviews

If a proposal is accepted, the guest editor will be responsible for:

- Preparing the “Call for Papers” to be included on the Journal’s Web site.
- Distribution of the Call for Papers broadly to various mailing lists and sites.
- Getting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Instructions for Authors.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

Special Issue for a Conference/Workshop

A special issue for a Conference/Workshop is usually released in association with the committee members of the Conference/Workshop like general chairs and/or program chairs who are appointed as the Guest Editors of the Special Issue. Special Issue for a Conference/Workshop is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

Guest Editors are involved in the following steps in guest-editing a Special Issue based on a Conference/Workshop:

- Selecting a Title for the Special Issue, e.g. “Special Issue: Selected Best Papers of XYZ Conference”.
- Sending us a formal “Letter of Intent” for the Special Issue.
- Creating a “Call for Papers” for the Special Issue, posting it on the conference web site, and publicizing it to the conference attendees. Information about the Journal and Academy Publisher can be included in the Call for Papers.
- Establishing criteria for paper selection/rejections. The papers can be nominated based on multiple criteria, e.g. rank in review process plus the evaluation from the Session Chairs and the feedback from the Conference attendees.
- Selecting and inviting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Author Instructions. Usually, the Proceedings manuscripts should be expanded and enhanced.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.