**In the Name of God**

# Journal of
## Information Systems & Telecommunication
**Vol. 1, No. 4, October-December 2013**

**Research Institute for Information and Communication Technology**
**Iranian Association of Information and Communication Technology**

**Affiliated to: Academic Center for Education, Culture and Research (ACECR)**


**Publisher**: Iran University Press
**Print ISSN**: 2322-1437
**Online ISSN**: 2345-2773
**Publication license:** 91/13216

**Editorial Office Address:** No.5, Saeedi Alley, Kalej Intersection., Enghelab Ave., Tehran, Iran,
P.O.Box: 13145-799 Tel: (+9821) 88930150 Fax: (+9821) 88930157
Email: info@jist.ir
URL: www.jist.ir

## Indexed in:

- Research Institute for Information and Communication Technology www.ictrc.ir
- Islamic World Science Citation Center (ISC) www.isc.gov.ir
- Scientific Information Database (SID) www.sid.ir
- Regional Information Center for Sciences and Technology (RICeST) www.srlst.com
- Magiran www.magiran.com

This Journal is published with scientific supported by the
Advanced Information systems Research Group (AIS)

## Acknowledgement

JIST Editorial-Board would like to gratefully appreciate the following distinguished referees for spending their invaluable time and expertise in reviewing the manuscripts and their constructive suggestions, which had a great impact on the enhancement of this issue of the JIST Journal.

**(A-Z)**

- Abbas Ebrahimi Moghadam, Ferdowsi University of Mashhad, Mashhad, Iran
- Fatemeh Geran Gharakhili, Shahid Rajaee Training Teacher, Tehran, Iran
- Hossein Kamrani, Shiraz University, Shiraz, Iran
- Roghieh Karimzadeh Baee, ITRC (Iran Telecommunication Research Center), Tehran, Iran
- Hengameh Keshavarz, University of Sistan & Baluchestan, Zahedan, Iran
- Mohammad Khalaj-Amirhosseini, Iran University Science and Technology, Tehran, Iran
- Negin Ma'navizadeh, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Kamal Mohamed pour, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Sajad Mohammadzadeh, University Birjand, Birjand, Iran
- Gholamreza Moradi, Amirkabir University of Technology, Tehran, Iran
- Vahid Reza Nafisi, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran
- Hossein Nezamabadipour, Shahid Bahonar University of Kerman, Kerman, Iran
- Mojtaba Shirazi, Amirkabir University of Technology, University Central of Florida
- Shahriar Shirvani Moghadam, Shahid Rajaee Training Teacher, Tehran, Iran
- Javad Soleiman Meiguni, Khaje Nasir-edin Toosi University of Technology, Tehran, Iran
- Hadi Soltanizadeh, Semnan University, Semnan, Iran
- Farzad Tavakoli Hamedani, Semnan University, Semnan, Iran

# Preface

Dear Valued Readers

An essential part of high quality scientific research is peer review, which in turn relies upon clear and efficient sharing of scientific data between different scientific communities. In this year, the focus of this publication and its contributors has been on providing the most recent scientific research studies on telecommunication and information systems.

The current issue concludes the fourth number of first volume of Journal of Information Systems and Telecommunication (JIST). A comparison between the numbers of submitted articles in these four issues shows a moderate increase from 30 to 150 papers. Published articles in this issue have been authored by 17 persons from 6 Universities. The paper acceptance rate in this year is about 20%. More than 168 distinguished referees did an excellent job to select the best papers for publication and their expertise has had a great impact on the enhancement of the published articles. We would like to congratulate and appreciate all these people for their contributions in this success.

Our community has grown; thanks to the devotion and tireless efforts shown by the staff and contributors of the "JIST and we hope that in the coming year we can repay their dedication by providing our readers with a high quality accessible and above all beneficial magazine. Special thanks are also due to the editorial board of JIST who have contributed their time and efforts to make this magazine what it is today. All of us will continue to do our utmost to make the "JIST" a name recognized for excellence in the scientific community.

As the last noteworthy point in this issue, we would like to attract the attention of our colleagues to the fifth issue. Keeping in mind that the deadline of paper submission is February 15, 2014, the interested people are kindly invited to read the call for papers page in www.jist.ir.

Fine Regards,
Masoud Shafiee
Editor- in- Chief

# Table of Contents

# Multiple Antenna Relay Beamforming for Wireless Peer to Peer Communications

Mohammad Hossein Golbon Haghighi*
Electrical Engineering - Communication, K. N. Toosi University of Technology, Tehran, Iran
mh.golbon@ee.kntu.ac.ir
Behrad Mahboobi
Electrical Engineering - Communication, K. N. Toosi University of Technology, Tehran, Iran
b.mahboobig@ee.kntu.ac.ir
Mehrdad Ardebilipour
Electrical Engineering - Communication, K. N. Toosi University of Technology, Tehran, Iran
mehrdad@eetd.kntu.ac.ir

Received: 14/Sep/2013 Accepted: 11/Dec/2013

**Abstract**
This paper deals with optimal beamforming in wireless multiple-input-multiple-output (MIMO) relay networks that involves multiple concurrent source-destination pairs with imperfect channel state information (CSI) at the relays. Our aim is the optimization of the MIMO relay weights that minimize the total relay transmit power subject to signal-to-interference-plus-noise ratio (SINR) of all destinations to be kept above a certain threshold. Since power minimization is a non-convex quadratically constrained quadratic programming (QCQP), we use semi-definite programming (SDP) relaxation of above mentioned problem by using a randomization technique. Numerical Monte Carlo simulations verify the performance gain of our proposed multiple antenna relay system in terms of transmit power and symbol error probability.

**Keywords:** MIMO-Relay Networks, Power Allocation, Beamforming , Semi-Definite Programming

## 1. Introduction

Recently, using MIMO-relays in wireless networks has attracted significant attention. Due to the shadowing, multipath fading and interference, the link quality between the source and destination in a wireless network degrades intensively. Several schemes to achieve spatial diversity are considered in literature. The most popular cooperative schemes are amplify-and-forward (AF) [1], decode-and-forward (DF)[2], compress-and forward [3] and coded cooperation [4]. In the AF scheme, sources transmit messages to the relays, which then simply scale their received signals according to a power constraint and forward the scaled signals toward the destinations. AF scheme has received extensive attention due to its simplicity.

In [5], distributed beamforming relay system with single transmitter-receiver pair and several relaying nodes has been proposed and perfect CSI knowledge is supposed to be available for each node. The authors in [6] have investigated the same scenario as in [5], but have assumed that the second-order statistics of all channel coefficients are available at the receiver. In [6,7], the beamforming weights are obtained in order to maximize the signal-to-noise ratio (SNR) at the destination subject to individual relay power. In the subsequent scenario, the problem has been solved subject to the total relay power constraints, while [5] solved the beamforming weights only subject to individual relay power constraints. Although, both of them have the same problem formulation, they have completely different approaches to solving the problem.

MIMO systems attracted considerable attention because of their ability to support high data rate and wireless network improvement [8]. In this paper, we aim to design optimal beamforming in MIMO relay networks.

Conic optimization techniques, as a result of modern convex optimization, have been extensively used in [9] to obtain a computational attractive problem emerged from the original difficult problem. The optimization problem of [7,10,11] is shown to be nonconvex quadratically constrained quadratic program, which can be solved by relaxing the original problem SDP problem [12], and employing the interior point methods (IPM) [13] for solving the SDP problem. Its problem has been solved efficiently because their solutions have been rank-one.

The aim of this paper is the optimization of MIMO relay weights that minimize the total relay transmit power subject to SINR of all destinations be kept above a certain threshold. We show that such a power minimization problem is a non-convex QCQP problem. We turn it into a semi-definite programing problem using a well-known relaxation technique which is NP-hard in general, but in our case can be efficiently and exactly solved using a randomization technique.

* Corresponding Author

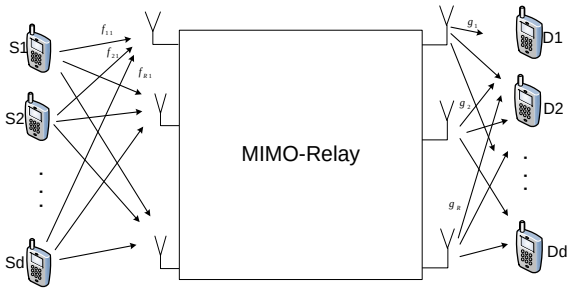



Figure 1: A MIMO-Relay Multiuser Network

## 2. System Model

We consider a peer to peer MIMO-relay network with d pairs of source-destination nodes as shown in Fig. 1. We have assumed that all source and destination nodes are equipped with one antenna and each source aims to maintain communication with its corresponding destination. All nodes work in half duplex mode and it is assumed that there is no direct link between source and destination. We use a two-step AF protocol. During the first step, each source broadcasts its signals to MIMO-relay. For the $P_{th}$ user, given $s_p$ as the source signal, the received signals at the MIMO-relay are collectively given as

$$\chi_r = \sum_{p=1}^{d} f_{rp} s_p + v'_r \in \mathbb{C}^{R\times 1}$$

where $f_{rp}$ is the channel coefficient from the $P_{th}$ source to the $r_{th}$ MIMO-relay antenna, $s_p$ is the information symbol assumed to be uncorrelated i.e.$E\{s_p s_q^*\} = P_p \delta_{pq}$ and $v'_r$ represents the AWGN at the MIMO relay, whose components are i.i.d. zero mean circularly symmetric Gaussian noise with unit variance ,i.e., CN(0,1). For simplicity, we rewrite the vector notation of the received signals as

$$x = \sum_{p=1}^{d} \mathrm{f}_p s_p + \mathrm{V}' \quad (1)$$

where

$$X = [X_1, X_2, \dots, X_R]^T, V' = [v_1, v_2, \dots, v_R]^T$$

and $\mathrm{f}_p = \left[f_{1p}, f_{2p}, \dots, f_{Rp}\right]^T$

At MIMO-relay, the received signal for $P_{th}$ user is processed by a complex beamforming weights $\mathbf{W} \in \mathbb{C}^{R\times R}$, which should be designed appropriately. During the 2nd step, the adjusted signal retransmitted by MIMO-relay is

$$\mathbf{t} = \mathbf{W}^H x \quad (2)$$

where $\mathbf{t}$ is an $R \times 1$ vector whose $r_{th}$ entry is the signal transmitted by the $r_{th}$ MIMO-relay antenna. Let us denote the vector of the channel coefficient from the MIMO-relay to the $k_{th}$ destination as $\boldsymbol{g}_k = [g_{1k}\ g_{2k} \dots\ g_{Rk}]^T$ .The received signal at the $k_{th}$ destination is expressed as

$$y_k = \mathbf{g}_k^T \mathbf{t} + n_k = \mathbf{g}_k^T \mathbf{W}^H \boldsymbol{x} + n_k$$

$$= \mathbf{g}_k^T \mathbf{W}^H \sum_{p=1}^{d} \mathbf{f}_{rp} \boldsymbol{s}_p + \mathbf{g}_k^T \mathbf{W}^H \mathbf{V}' + n_k$$

$$= \underbrace{\mathbf{g}_k^T \mathbf{W}^H \mathbf{f}_k s_k}_{\text{desired signal}} + \underbrace{\mathbf{g}_k^T \mathbf{W}^H \sum_{p=1,p=k}^{d} \mathbf{f}_{rp} \boldsymbol{s}_p}_{\text{Interfereence}} + \underbrace{\mathbf{g}_k^T \mathbf{W}^H \mathbf{V}' + n_k}_{\text{totalnoise}} \quad (3)$$

where $n_k$ is the noise at the receiver, which is also assumed to be CN(0,1).

## 3. Optimal Relay Power Control

We aim to find the beamforming weights such that the MIMO-relay transmit power is minimized while maintaining the destinations QOS at a certain level, i.e., every destination SINR is required to be larger than a certain threshold value. The optimization problem is formulated as follows:

$$\underset{w}{\text{Minimize}}\ P_T$$
$$\text{Subject to}\ SINR_k \geq \gamma_k,\ for\ k = 1, 2, \dots, d \quad (4)$$

where $P_T$ is the MIMO-relay transmit power, $SINR_k$ and $\gamma_k$ are the SINR and the target SINR at the $k_{th}$ destination, respectively. Then, the SINR at the $k_{th}$destination is given by

$$SINR_k = \frac{P_s^k}{P_n^k + P_i^k} \quad (5)$$

Here, $P_s^k$ , $P_n^k$ and $P_i^k$ represent the power of desired signal, noise and interference at the $k_{th}$ destination, respectively.

We now derive the expressions for the total transmit power of the MIMO-relay and $SINR_k$. Using (2), the total MIMO-relay transmit power can be obtained as

$$P_T = E\{\mathbf{t}^H \mathbf{t}\} = E\{X^H \mathbf{W}\mathbf{W}^H X\} \quad (6)$$
$$= \mathrm{trace}\{\mathbf{W}^H \mathbf{R}_x \mathbf{W}\}$$

where $\mathbf{R}_x \triangleq E\{XX^H\}$ and can be expressed $\mathbf{R}_x$ as

$$\mathbf{R}_x = \sum_{P=1}^{d} P_P\, E\{\mathbf{f}_p \mathbf{f}_p^H\} + \sigma_v^2 \mathbf{I}_{R\times R} \quad (7)$$

where $\mathbf{R}_f^p \triangleq E\{\mathbf{f}_p \mathbf{f}_p^H\}$ and trace{.} represents the trace of a matrix.

Using Kronecker identity, we have

$$trace(\mathbf{A}\mathbf{X}^H\mathbf{B}\mathbf{X}) = Vec(\mathbf{X})^H(\mathbf{A}^T \otimes \mathbf{B})Vec(\mathbf{X})$$

Then, we can rewrite the total transmit power of MIMO-Relay as

$$P_T = Vec(\mathbf{W})^H(\mathbf{I}_{R\times R} \otimes \mathbf{R}_x)Vec(\mathbf{W}) \quad (8)$$
$$= \mathbf{w}^H \mathbf{D}\mathbf{W}$$

Let us define $\mathbf{w} \triangleq Vec(\mathbf{W}) and \mathbf{D} \triangleq (\mathbf{I}_{R\times R} \otimes \mathbf{R}_x)$ where $Vec(.)$ is the vectorization operator which stacks all columns of a matrix on top of each other and $\otimes$

represents the kronecker multiplication of two matrices. Using desired signal component of (3), we can obtain $P_s^k$ as

$$\begin{aligned} P_s^k &= E\left\{\mathbf{g}_k^T \mathbf{W}^H \mathbf{f}_k \mathbf{f}_k^H \mathbf{W} \mathbf{g}_k^*\right\} E\left\{\left|s_k\right|^2\right\} \\ &= Vec(\mathbf{W})^H \left(\mathbf{R}_{\mathbf{g}_k}^T \otimes \mathbf{R}_{f_k}\right) Vec(\mathbf{W}) \\ &= \mathbf{w}^H \mathbf{R}_k \mathbf{w} \end{aligned} \quad (9)$$

where, $\mathbf{R}_{gk} \square E\{\mathbf{g}_k^* \mathbf{g}_k^T\}, \mathbf{R}_k \square (\mathbf{R}_{\mathbf{gk}}^T \otimes \mathbf{R}_{\mathbf{f}k})$, and the total noise power is given by

$$\begin{aligned} P_n^k &= E\left\{v^H \mathbf{W} \mathbf{g}_k^* \mathbf{g}_k^T \mathbf{W}^H v\right\} + \sigma_n^2 \\ &= \sigma_v^2 \text{trace}\left\{\mathbf{W}^H \mathbf{W} \mathbf{R}_{\mathbf{g}_k}\right\} + \sigma_n^2 \\ &= \sigma_v^2 \left\{Vec(\mathbf{W})^H \left(\mathbf{R}_{\mathbf{g}_k}^T \otimes \mathbf{I}\right) Vec(\mathbf{W})\right\} + \sigma_n^2 \\ &= \mathbf{w}^H \mathbf{D}_k \mathbf{w} + \sigma_n^2 \end{aligned} \quad (10)$$

where $\mathbf{D}_k = \sigma_v^2 (\mathbf{R}_{\mathbf{gt}}^T \otimes \mathbf{I})$.

Denoting $\mathbf{D}_k = \{1,2,\ldots,d\} - \{k\}$ and using the interference component of (3), the interference power at $k_{th}$ destination can be obtained as

$$\begin{aligned} P_i^k &= E\left\{\mathbf{g}_k^T \mathbf{W}^H \left(\sum_{p,q \in D_k} \mathbf{f}_p \mathbf{f}_q^H s_p s_q^*\right) \mathbf{W} \mathbf{g}_k^*\right\} \\ &= trace\left\{\mathbf{W}^H \left(P_p \sum_{p \in \mathbf{D}_k} \mathbf{R}_{\mathbf{f}_p}\right) \mathbf{W} E\left\{\mathbf{g}_k^* \mathbf{g}_k^T\right\}\right\} \\ &= trace\left\{\mathbf{W}^H \mathbf{A}_k \mathbf{W} \mathbf{R}_{\mathbf{g}_k}\right\} = Vec(\mathbf{W})^H \left(\mathbf{R}_{\mathbf{g}_k}^T \otimes \mathbf{A}_k\right) Vec(\mathbf{W}) \\ &= \mathbf{w}^H \mathbf{Q}_k \mathbf{w} \end{aligned} \quad (11)$$

where

$$\mathbf{A}_k \square \left(P_P \sum_{p=D_k} \mathbf{R}_{\mathbf{f}_p}\right), \mathbf{Q}_k \square (\mathbf{R}_{\mathbf{gt}}^T \otimes \mathbf{A}_k)$$

Finally, the optimization problem can be written as:

$$\begin{aligned} &\underset{w}{\text{Minimize}}\ \mathbf{w}^H \mathbf{D} \mathbf{w} \\ &\text{Subject to}\ \frac{\mathbf{w}^H \mathbf{R}_k \mathbf{w}}{\mathbf{w}^H \left(\mathbf{D}_k + \mathbf{Q}_K\right) \mathbf{w} + \sigma_n^2} \geq \gamma_k \\ &\qquad for\ k = 1,2,\ldots,d \end{aligned} \quad (12)$$

Since $\mathbf{w}^H (\mathbf{D}_k + \mathbf{Q}_K) \mathbf{w} + \sigma_n^2 \geq 0$, the above problem is equivalent to

$$\begin{aligned} &\underset{w}{\text{Minimize}}\ \mathbf{w}^H \mathbf{D} \mathbf{w} \\ &\text{Subject to}\ \mathbf{w}^H \left(\mathbf{R}_k - \gamma_k \left(\mathbf{D}_k + \mathbf{Q}_K\right)\right) \mathbf{w} \geq \gamma_k \sigma_n^2 \\ &\qquad for\ k = 1,2,\ldots,d \end{aligned} \quad (13)$$

Let us define $\mathbf{T}_K \square \mathbf{R}_k - \gamma_k (\mathbf{D}_k - \mathbf{Q}_k)$. In our QCQP problem, since all the matrices $\mathbf{T}_k$ are negative semidefinite for all k, the problem is convex and can be solved efficiently. However, the feasible set of our optimization problem is empty, since $\mathbf{w}^H \mathbf{T}_k \mathbf{w} \leq 0$ for all *k* and *w*. Therefore, the non-convex equality constraint (13) reveals that the QCQP problem is non-convex and NP-hard in general. However, we will show that an exact and simple solution can be found in our specific problem

## 4. Solution via SDP Relaxation

We first turned our QCQP problem into the semidefinite programing (SDP) problem. Let us define $\mathbf{X} \square \mathbf{w}\mathbf{w}^H$, thus we recast the problem as follows:

$$\begin{aligned} &\underset{X}{\text{Minimize}}\ \text{trace}\left(\mathbf{D}\mathbf{X}\right) \\ &\text{Subject to}\ \text{trace}\left(\mathbf{T_k}\mathbf{X}\right) \geq \gamma_k \sigma_n^2,\ for\ k = 1,2,\ldots,d \\ &\qquad and\ Rank(\mathbf{X}) = 1, \mathbf{X} \geq 0 \end{aligned} \quad (14)$$

The rank-one constraint on **X** is nonconvex, hence the problem is nonconvex. By dropping the nonconvex rank-one constraint, the problem can be relaxed to a convex SDP problem. The relaxed version of the problem **Error! Reference source not found.** can be represented by the following SDR[1] form. Now we can obtain a lower bound on the optimal value of **Error! Reference source not found.** by solving this relaxed problem.

$$\begin{aligned} &\underset{\mathbf{X}}{\text{Minimize}}\ \text{trace}\left(\mathbf{D}\mathbf{X}\right) \\ &\text{Subject to}\ \text{trace}\left(\mathbf{T_k}\mathbf{X}\right) \geq \gamma_k \sigma_{\varsigma_k}^2, \text{for}\ k = 1,2,\ldots,d \\ &\qquad \mathbf{X} \geq 0 \end{aligned} \quad (15)$$

Modern SDP solvers, such as SeDuMi [14,15], use interior point methods to find an efficient optimal solution for the problem, if it be feasible; otherwise, they return to an assertion of infeasibility. Generally, the optimal value of SDP problem is a lower bound for the optimal value of the nonconvex QCQP problem, because the feasible set of problem **Error! Reference source not found.** is only a subset of the feasible set of problem **Error! Reference source not found.**. If the optimal value in **Error! Reference source not found.** ,i.e. $\mathbf{X_{opt}}$, is rank-one, then its principal eigenvector is exactly the optimal solution of the original optimization problem, otherwise, a rank-one solution for the original problem can be found using a randomization technique. The idea behind this technique is to generate candidate sets of beamforming vectors from the optimal solution matrix $\mathbf{X_{opt}}$ of problem **Error! Reference source not found.** [16]. The accuracy of these techniques for semidefinite problem has been analyzed for different problems in [17,18] and it has been found that the randomization has acceptable performance in practical scenarios. In order to achieve this goal, first the eigenvalue decomposition of $\mathbf{X_{opt}}$ should be calculated as $\mathbf{X_{opt}} = VDV^H$. Then the candidate vector is generated as $x_c = VD^{(1/2)}P_c$, where $P_c$ is a circularly symmetric complex white Gaussian vector generated as $P_c \in \square^{R \times 1} \square Nc(0,1)$. Hence, it can be recognized that the vector $x_c$ satisfies $E\{x_c x_c^H\} = \mathbf{X_{opt}}$. This random vector generation procedure should be performed multiple times and in each iteration, any vector (or its scaled version) that satisfies SINR constraints of problem **Error! Reference source not found.** is saved as a candidate vector $x'_c$ along with corresponding objective values. The vector generation should be performed for a

[1] Semi Definite Representation

predetermined number of times. The final minimum solution can be obtained by a simple minimization over the finite set of objective values as an approximate solution for the problem **Error! Reference source not found.**. One way to solve the problem **Error! Reference source not found.** by $x_c$ is to find a proper scaling factor $\sqrt{\beta} \geq 0$. Applying $\beta$ to **Error! Reference source not found.**, the following problem will be attained

$$\begin{aligned} &\underset{\mathbf{X}}{\text{Minimize}}\ \beta\,\text{trace}(\mathbf{DX}) \\ &\text{Subject to}\ \beta\,\text{trace}(\mathbf{T_k X}) \geq \gamma_k \sigma_{\varsigma_k}^2\ \text{, for } k = 1,2,...,d \\ &\qquad \text{Rank}(\mathbf{X}) = 1\ \text{, } \mathbf{X} \geq 0 \end{aligned} \tag{16}$$

In the above algorithm, the acceptable scaling factors are those that $\beta\,\text{trace}(\mathbf{T}_k\mathbf{X}) \geq 0$. Thus, the maximum scaling factor should be selected as [10]:

$$\beta = \max_{k=1,...,d} \left\{ \frac{\gamma_k \sigma_{\varsigma_k}^2}{\text{trace}(\mathbf{T}_k \mathbf{X})} \right\} \tag{17}$$

Consequently, the approximate solution of problem **Error! Reference source not found.** is $\sqrt{\beta} x_c$. In our case, after an acceptable number of iterations (around 40 iterations), the solution of the randomization problem approached to its lower bound (the optimal value of relaxed problem). Therefore, $\mathbf{X_{opt}}$ is an acceptable and near optimal solution for the original nonconvex problem.

Another Suboptimal solution of the original problem **Error! Reference source not found.** can be found by using penalty function in the objective part of the problem and converting the objective function into the difference of two convex functions [19] subject to current convex constraints. [20,21] have developed an effective nonsmooth optimization algorithm based on the sub-gradient of Rank one function.

## 5. Computational Complexity

The aim of this section is to analyze the computational complexity of related the MIMO relay systems used in practice. According to records, Nesterov and Nemirovskii [22] were the pioneers who extended interior-point methods from linear optimization to semi-definite optimization problem and built up the polynomial complexity of the algorithm, at least in theory.

In this part, we assess the computational complexity of a standard problem with only equality constraints and then we extend conclusions to our case with inequality and/or equality constraints. The standard form of SDP problem is defined as:

$$\begin{aligned} &\underset{\mathbf{X}}{\text{Minimize}}\ \ \text{trace}(\mathbf{CX}) \\ &\text{Subject to}\ \ \text{trace}(\mathbf{A}_i \mathbf{X}) = b_i\text{, for i} = 1,...,\text{d} \\ &\qquad \mathbf{X} \geq 0 \end{aligned} \tag{18}$$

where $\mathbf{C}$ and $\mathbf{A}_i (\text{i} \leq \text{i} \leq \text{d})$ are symmetric $n \times n$ matrices and $\text{b} \in R^d$. Thus, for such a problem the complexity order with large-update (or long-step) algorithm based on the primal dual SDP algorithm is

$$\text{O}\left(\sqrt{n} \log n \log\left(n/\varepsilon\right)\right) \tag{19}$$

where $\varepsilon$ denotes the accuracy parameter of the algorithm, while this algorithm with small-update (or short-step) still has $\text{O}(\sqrt{n}\, log(n/\varepsilon))$ iterations bound [13].

Small-update IPMs are confined to unacceptably slow progress, while large-update IPMs are more efficient for faster progress in practice. Although, large-update IPMs perform much more efficiently and IPM algorithms are effective in practice, they often have somewhat worse complexity bounds.

The complexity order of solving the SDP problem in **Error! Reference source not found.** is polynomial time. To evaluate the complexity of the original problem, the dimension parameter n should be specified.

Therefore, the dimensions of the matrices used in the objective and constraints of the problem **Error! Reference source not found.** should be determined.

Regarding to Kronecker product of two matrices, if $\mathbf{A} \in \square^{n \times n}$ and $\mathbf{B} \in \square^{m \times m}$, then $\mathbf{A} \otimes \mathbf{B}$ will be a $nm \times nm$ matrix. According to the vectors definite in (8) and the sizes of $\mathbf{R}_x \in \square^{R \times R}$ and $\mathbf{I} \in \square^{R \times R}$, the dimension of $\mathbf{D}$ will be

$$\mathbf{D} \in \square^{R^2 \times R^2}$$

The same approach can be used to find the size of $\mathbf{X}$ and $\mathbf{T}_k$, and as a result we have: $\mathbf{X}, \mathbf{T}_k \in \square^{R^2 \times R^2}$. Consequently, the dimension of matrices $\mathbf{X}, \mathbf{T}_k, \mathbf{D}$ is $R^2$. It is notable that the constraints of our problem are not the same as the ones in the standard SDP problem. Therefore, they have to be modified to be similar to the standard format. In order to achieve this goal, the first step is to define $y_i$ so that the inequality constraints of **Error! Reference source not found.** change to equality relations.

$$\begin{aligned} &\text{trace}(\mathbf{T}_\text{i}\mathbf{X}) = \gamma_\text{i}\sigma_{\varsigma_k}^2 + y_\text{i}\text{, for i} = 1,...,\text{d} \\ &\mathbf{X} \geq 0, y_\text{i} \geq 0 \end{aligned}$$

Next, a new variable $\hat{\mathbf{X}}$ should be defined in order to standardize the problem:

$$\hat{\mathbf{X}} \square \begin{bmatrix} \mathbf{X} & \mathbf{0}_{R^2 \times d} \\ \mathbf{0}_{d \times R^2} & \begin{bmatrix} y_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & y_d \end{bmatrix} \end{bmatrix}$$

As a result, the following standard form will be attained.

$$\begin{aligned} &\underset{\mathbf{X}}{\text{Minimize}}\ \ \text{trace}\left(\hat{\mathbf{D}}\hat{\mathbf{X}}\right) \\ &\text{Subject to}\ \ \text{trace}\left(\hat{\mathbf{T}}_\text{i}\hat{\mathbf{X}}\right) = b_\text{i}\text{, for i} = 1,...,\text{d} \\ &\qquad \hat{\mathbf{X}} \geq 0 \\ &\text{where } \hat{\mathbf{D}} \square \begin{bmatrix} \mathbf{D} & \mathbf{0}_{R^2 \times d} \\ \mathbf{0}_{d \times R^2} & \mathbf{0}_{d \times d} \end{bmatrix}, \hat{\mathbf{T}}_i \square \begin{bmatrix} \mathbf{T}_i & \mathbf{0}_{R^2 \times d} \\ \mathbf{0}_{d \times R^2} & \mathbf{0}_{d \times d} \end{bmatrix} \end{aligned} \tag{20}$$

Then, it is clear that

$$n = R^2 + d \square R^2 \quad (21)$$

Consequently, the worst case complexity of solving **Error! Reference source not found.** is

$$O(R\log(R^2)\log(R^2/\varepsilon)) \quad (22)$$

## 6. Numerical Results

In this section, we examine the performance of the proposed MIMO relay system in various scenarios. The channel vectors, **f** and **g** are assumed to be statistically independent and generated as i.i.d. complex Gaussian random variables with zero mean and different values of variances. Moreover, in all simulation results, the output power of all sources are assumed to be equal, i.e., $\{P_k\}_{k=1}^{d} = P = 0$dB, and also

$$\{\gamma_k\}_{k=1}^{d} = \gamma_{th} \quad , \quad \{\sigma_v^2\}_{i=1}^{R} = \{\sigma_{\varsigma_k}^2\}_{k=1}^{d} = \sigma^2$$

Throughout our numerical examples, the noise power is normalized by the source transmit power, i.e. $\sigma^2 = 0dB$. Also, we assume that all of the channel coefficients are exactly known at a processing center in which the beamforming weights should be optimized.

In all of the cases, we solved the relaxed version of optimization problem using CVX software. If $\mathbf{X_{opt}}$ is rank-one matrix, its principal component is used to determine the solution of problem **Error! Reference source not found.**; otherwise the best rank-one approximation is obtained using the randomization procedure as described in section 4. In all of our scenarios, the relay transmit power is plotted for those values which qualify QoS constraint and each curve is plotted only for those threshold values for which the beamforming problem is feasible for at least 80% of the total realizations.

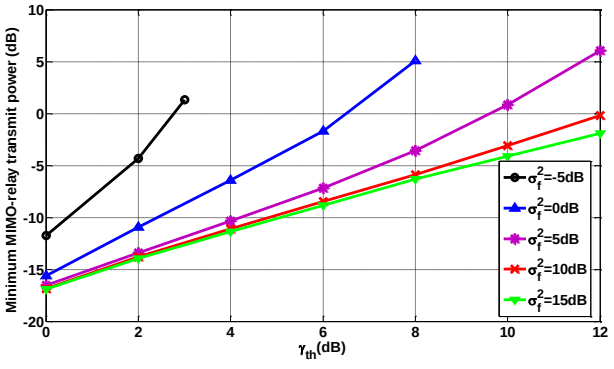



Figure 2: Minimum MIMO-relay transmit power $P_T^{min}$ versus destination SINR threshold value $\gamma_{th}$, for different values of $\sigma_f^2$ and $\sigma_g^2 = 10$dB

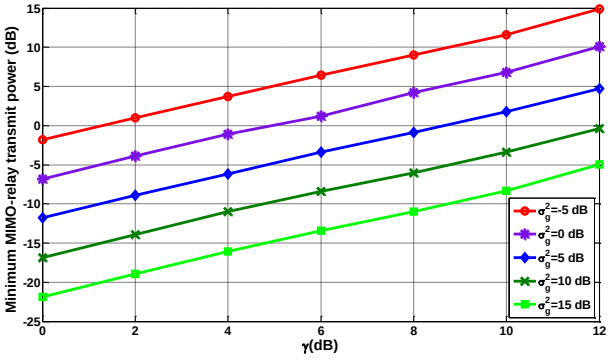



Figure 3: Minimum MIMO-relay transmit power $P_T^{min}$ versus destination SINR threshold value $\gamma_{th}$, for different values of $\sigma_g^2$ and $\sigma_f^2 = 10$dB

$\sigma_g^2$ and $\sigma_f^2$ show the effect of the quality of the downlink and uplink channels, respectively. In Figure 2, we have plotted minimum MIMO-relay transmit power $P_T^{min}$ versus destination SINR threshold value $\gamma_{th}$, for $\sigma_g^2 = 10dB$ and different values of $\sigma_f^2$. In this simulation setup, we have quantified the total relay transmit power by changing the SINR threshold value $\gamma_{th}$, from 0 dB to 25dB. The results are averaged over 1000 realizations of channel coefficients with CSG [1] distribution shown in Figure 2 for different values of $\sigma_f^2$. As can be seen from this figure, the better quality of uplink channel, the less minimum transmit power required to meet a certain QOS.

Moreover, as expected, this figure shows that the transmit power for all cases increases with increasing $\gamma_{th}$, due to the fact that, in order to achieve higher QOS requirement, the MIMO relay needs to spend more of its transmit power.

Figure 3 illustrates the minimum transmit power of MIMO-relay versus $\gamma_{th}$, for $\sigma_f^2 = 10dB$ and different values of $\sigma_g^2$. This figure also shows the effect of the quality of downlink channel on the minimum transmit power.

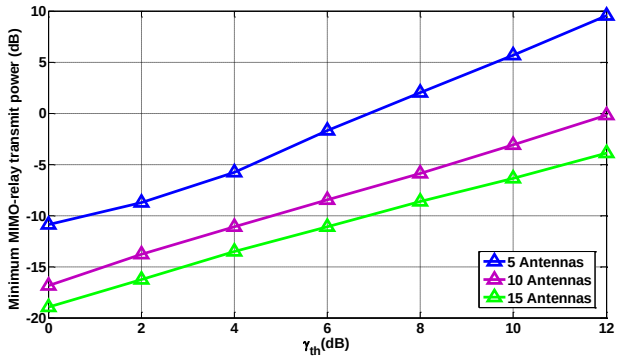



Figure 4: Minimum MIMO relay transmit power $P_T^{min}$ versus destination SINR threshold value $\gamma_{th}$, for different number of antennas.

[1] Complex Symmetric Gaussian

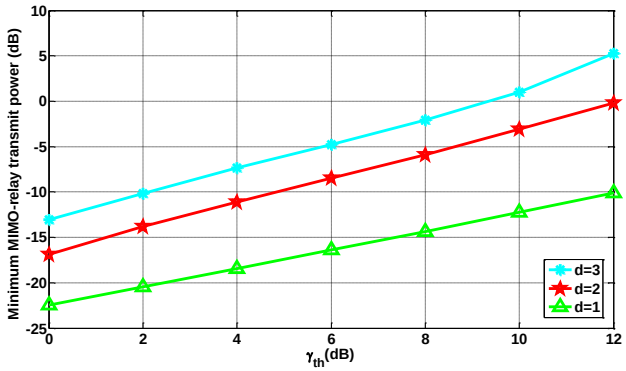



Figure 5: Minimum MIMO relay transmit power $P_T^{min}$ versus destination SINR threshold value $\gamma_{th}$, for different number of source-destination pairs.

To study the effect of the number of MIMO-relay antennas in terms of quality of channels, we consider three different networks with different number of relay antennas, while setting $\sigma_f^2 = 10dB$, $\sigma_g^2 = 10dB$. Figure 4 illustrates the average minimum transmit power of MIMO-relay versus $\gamma_{th}$ for two users and different number of antennas. As expected, it is observed that for a certain minimum value of received SINR, more power saving will be obtained by increasing the number of antennas. It can be seen that increasing the number of antennas from 5 to 10 results in at least 9 dB improvement in transmit power. This performance improvement decreases as the number of antennas increases.

In Figure 5, we investigate the performance of the network by changing the number of source-destination pairs. This figure reveals that the lower the number of user, the lower the minimum MIMO-relay transmit power required to meet a certain threshold level.

## 7. Conclusions

This paper has studied the optimal beamforming design in wireless multi-user MIMO-relay network to minimize MIMO-relay transmit power with guaranteed QOS at destinations. The proposed designs are based on a two-step amplify-and-forward protocol and imperfect channel state information. It has been shown that the corresponding optimization problem is nonconvex, but it can be converted into a convex problem by using a semidefinite relaxation technique and can be solved efficiently and accurately using well-known randomization technique.

## References


[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062-3080, Dec. 2004.

[2] J. N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415-2425, Oct. 2003.

[3] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3037-3063, Sep.2005.

[4] A. N. Todd and E. Hunter, "Diversity through Coded Cooperation," *IEEE Trans. Wireless Commun*., vol. 2, pp. 283-289, Feb 2006.

[5] J. Yindi and H. Jafarkhani, "Network Beamforming Using Relays With Perfect Channel Information," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2499-2517, Jun 2009.

[6] V. Havary-Nassab, S. Shahbazpanahi, A. Grami, and L. Zhi-Quan, "Distributed Beamforming for Relay Networks Based on Second-Order Statistics of the Channel State Information," *IEEE Trans. Signal Process*., vol. 56, pp. 4306-4316, Sep. 2008.

[7] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Gazor, "Multiple Peer-to-Peer Communications Using a Network of Relays," *IEEE Trans. Signal Processing*, vol. 57, pp. 3053-3062, Aug. 2009.

[8] A. S. Behbahani, R. Merched, and A. M. Eltawil, "Optimizations of a MIMO Relay Network," *IEEE Trans. Signal Process*., vol. 56, pp. 5062-5073, Oct. 2008.

[9] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process*., vol. 54, pp. 2239-2251, Jun. 2006.

[10] B. K. Chalise and L. Vandendorpe, "MIMO Relay Design for Multipoint-to-Multipoint Communications With Imperfect Channel State Information," *IEEE Trans. Signal Process*., vol. 57, pp. 2785-2796, Jul 2009.

[11] B. K. Chalise and L. Vandendorpe, "Joint optimization of multiple MIMO relays for multi-point to multi-point communication in wireless networks," in *Signal Processing Advances in Wireless Communications, 2009. SPAWC '09. IEEE 10th Workshop on, 2009*, pp. 479-483.

[12] A. d'Aspremont and S. Boyd. *Relaxations and Randomized Methods for Nonconvex QCQPs. Stanford, CA: Stanford Univ., Autumn 2003 [Online]. Available: www.standford.edu/class/ee392o, EE92o.*

[13] J. Peng, C. Roos, and T. Terlaky, *Self-Regularity: A New Paradigm for Primal-Dual Interior-Point Algorithms*: Princeton University Press, 2002.

[14] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimiz. Methods Softw*., vol. 11-12, pp. 625-653, 1999.

[15] M. G. a. S. Boyd. (2008). *CVX: Matlab Software for Disciplined Convex Programming*. Available: http://stanford.edu/boyd/cvx

[16] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of Service and Max-Min Fair Transmit Beamforming to Multiple Cochannel Multicast Groups," *IEEE Trans. Signal Process*., vol. 56, pp. 1268-1279, Mar. 2008.

[17] Z.-Q. Luo, N. D. Sidiropoulos, P. Tseng, and S. Zhang, "Approximation bounds for quadratic optimization with homogeneous quadratic constraints," *SIAM J. Optim*., vol. 18, pp. 1–28, Feb. 2007.
[18] T.-H. Chang, Z.-Q. Luo, and C.-Y. Chi, "Approximation Bounds for Semidefinite Relaxation of Max-Min-Fair Multicast Transmit Beamforming Problem," *IEEE Trans. Signal Process*., vol. 56, pp. 3932-3943, Aug. 2008.
[19] H.Tuy, "Convex Analysis and Global Optimization," *Kluwer Academic*, 1998.
[20] A. H. Phan, H. D. Tuan, H. H. Kha, and H. H. Nguyen, "Beamforming Optimization in MultiUser Amplify-and-Forward Wireless Relay Networks," *IEEE Trans. Wireless Commun*., vol. 11, pp. 1510-1520, Apr. 2012.
[21] U. Rashid, H. D. Tuan, and H. H. Nguyen, "Maximin relay beamforming in multiuser amplify-forward wireless relay networks," presented at the Wireless Commun. and Networking Conf. (WCNC), 2012.
[22] Y. E. Nesterov and A. S. Nemirovskii, " Interior point polynomial algorithms in convex programming," S*IAM Studies in Applied Mathematics*, vol. 13,SIAM ,Philadelphia,PA, 1994.

**Mohammad Hossein Golbon Haghighi** received the BSc. Degree in electrical engineering from Shiraz University of Technology in 2009 and MSc degree in communication systems from K. N. Toosi university of Technology in 2011. His research interests include Wireless Network Resource Allocation; Multiuser Multi-relay Networks; CMDA in Wireless Relay Networks; Optimization methods in Communication Networks, Cognitive Radio Networks, and Routing in Wireless Networks.

**Behrad Mahboobi** received BSC. Degree in electrical engineering from K.N.Toosi University of technology, Tehran, Iran, in 2003 and his Master Degree from University of Tehran in 2006. He is Currently, Ph.D. Student at the University of K.N.Toosi in studying Telecommunication engineering. He worked in telecommunication industy as researcher to develop novel communication system and implement signal processing and communication algorithm in FPGA. He also participates in the development of next generation network protocol and security system. His major academic research interest includes Game theory, convex optimization, Signal Processing, Distributed computing, Cooperative networks, Information Theory and wireless networks.

**Mehrdad Ardebilipour** was born in Iran, in February 1954. He received B.Sc. and M.Sc. Degrees in electrical engineering from K. N. Toosi University of Technology, Tehran, Iran, in 1977 and Tarbiat Modarres University, Tehran, Iran, in 1989, respectively. He has also been awarded the degree of Ph.D. by the university of Surrey, Guildford, England, in 2001. Since 2001, he has been an assistant professor at K. N. Toosi University of Technology and was directing the Communications Engineering Department for 6 years. Currently, he is the director of the Spread Spectrum and Wireless Communications research laboratory. His current research interests are cognitive radio, cooperative communication, ad-hoc and sensor networks, MIMO communication, OFDM, game theory and cross-layer design for wireless communications.

# Parameter Estimation in Hysteretic Systems Based on Adaptive Least-Squares

Mansour Peimani*
Department of Electrical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
m.peimani@srbiau.ac.ir
Mohammad Javad Yazdanpanah
Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Iran
yazdan@ut.ac.ir
Naser Khaji
Faculty of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran
nkhaji@modares.ac.ir

Received: 12/Aug/2013 Accepted: 11/Dec/2013

**Abstract**
In this paper, various identification methods based on least-squares technique to estimate the unknown parameters of structural systems with hysteresis are investigated. The Bouc-Wen model is used to describe the behavior of hysteretic nonlinear systems. The adaptive versions are based on the fixed and variable forgetting factor and the optimized version is based on optimized adaptive coefficient matrix. Simulation results show the efficient performance of the proposed technique in identification and tracking of hysteretic structural system parameters compared with other least square based algorithms.

**Keywords:** Hysteresis, Least-Squares Estimation (LSE), Optimization, System Identification.

## 1. Introduction

The System identification and fault detection based on measured vibration data through condition monitoring systems has been noticeable in recent years. Identifying the status of a structure and fault detection is the main goal of the condition monitoring systems for civil structures. Different methods of data analysis are reviewed in [1], which includes frequency domain analysis and time domain analysis methods. The advantages and shortcomings of damage identification methods are analyzed in [2]. Identification of structural systems can be categorized into two parts: online and offline. For online structural parameters changes identification, time domain analysis methods such as least-squares estimation [3-5] and filter-based methods such as Kalman filter [6,7], $H_\infty$ filters [8], and wavelet technique [9] are used. Today, real-time detection of changes in structural parameters due to failures during events such as earthquakes is a challenging issue. Civil structures faced with intense earthquakes usually show hysteresis behavior. Various models have been proposed to identify and simulate the hysteresis, and Bouc-Wen model is the most appropriate [10]. This model is a quasi-physical and It can be used to describe the behavior of the wide range of considered systems [11,12].

Least-squares parameter estimation algorithm cannot estimate the time-varying parameters well. Adaptive LSE method to estimate time-varying parameters have been presented in [3,5]. A frequency domain nonlinear least-squares estimation algorithm was proposed in [13]. Fuzzy least-squares estimator was investigated in [14] by proposing a confidence region. A new structured total least-squares based frequency estimation algorithm for real sinusoids corrupted by white noise was adapted in [15].

In this paper, least-squares estimation algorithm and the adaptive versions based on the fixed and variable forgetting factor and optimized version to determine coefficients for tracking time-varying parameters are presented. Considered methods are applied for online identification of parameter changes of the nonlinear structural systems with hysteresis. The ability of the methods to track instantaneous changes in the parameters of a structural system due to failures is evaluated.

## 2. Problem Statement

Motion Equations of a *m* degrees of freedom can be described with Eq. (1).

$$M\ddot{x}(t) + F_c[\dot{x}(t)] + F_s[x(t)] = \eta f(t) \quad (1)$$

Where *M* is the mass matrix, *x(t)* is the displacement vector, $F_C[\dot{x}(t)]$ is the dissipative force vector, $F_s[x(t)]$ is the non dissipative restoring force vector, *f(t)* is the excitation vector, and $\eta$ is the excitation influence matrix. Suppose we have a structure for estimating the unknown parameters including damping, stiffness, and hysteresis parameters; i.e., $[\theta_1(t), \theta_2(t), \dots, \theta_n(t)]^T$. The observation equation associated with the motion equation of the structural system is expressed as

$$\varphi[\ddot{x}, \dot{x}, x; t]\theta(t) + \varepsilon(t) = y(t) \quad (2)$$

* Corresponding Author

where $\theta(t)$ is a vector of $n$ unknown parameters. $\ddot{x}$ , $\dot{x}$ and $x$ are the measured acceleration, velocity and displacement response vectors. $y(t)$ and $\varepsilon(t)$ are the excitation and noise vectors. $\varphi$ is a $m \times n$ data matrix. The Eq. (2) can be written at the time instant $t = t_k$ as

$$\varphi_k \theta_k + \varepsilon_k = y_k \tag{3}$$

The solution of the recursive LSE, $\hat{\theta}_{k+1}$ to estimate $\theta_{k+1}$ are given by the following equations [16].

$$\begin{aligned} &\hat{\theta}_{k+1} = \hat{\theta}_k + K_{k+1}[y_{k+1} - \varphi_{k+1}\hat{\theta}_k] \\ &K_{k+1} = P_k \varphi_{k+1}^T [I + \varphi_{k+1} P_k \varphi_{k+1}^T]^{-1} \\ &P_k = [I - K_k \varphi_k] P_{k-1} \end{aligned} \tag{4}$$

where $K_{k+1}$ and $P_k^T = P_k > 0$ are the estimation and adaptation gain matrices.

## 3. Least-squares Estimation Based Algorithms

The Least-squares parameter estimation algorithm can not estimate the time-varying parameters well. A simple method is to use a fixed forgetting factor. In this method, $P_k$ is replaced by $\bar{\lambda}^{-1} P_k$, where $\bar{\lambda} \in (0,1]$ is a constant forgetting factor. The ability to track parameters changes and the sensitivity to measurement noises increase in this method for a small forgetting factor. To enhance the performance a constant factor $\bar{\lambda}$ is replaced by $\bar{\lambda}_k$ in the variable forgetting factor method. Although in both methods when a parameter is changed other parameters are estimated with oscillations. To solve these problems and improve the real-time changes in parameter estimation the adaptive coefficient matrix method is proposed where each factor is adaptive to a specific changing parameter. The estimation error is corrected by a adaptive factor $\lambda_j(k+1)$ for estimating the parameter $\theta_j(k+1)$. Recursive solution for the vector of variable parameters is achieved as follows [5].

$$\begin{aligned} &\hat{\theta}_{k+1} = \hat{\theta}_k + K_{k+1}[y_{k+1} - \varphi_{k+1}\hat{\theta}_k] \\ &K_{k+1} = (\Lambda_{k+1} P_k \Lambda_{k+1}^T)\varphi_{k+1}^T [I + \varphi_{k+1}(\Lambda_{k+1} P_k \Lambda_{k+1}^T)\varphi_{k+1}^T]^{-1} \\ &P_k = [I - K_k \varphi_k](\Lambda_k P_{k-1} \Lambda_k^T), \quad k = 1,2,\ldots \end{aligned} \tag{5}$$

where $\Lambda_{k+1} = diag[\lambda_1(k+1), \lambda_2(k+1), \ldots, \lambda_n(k+1)]$ is a diagonal adaptive factor matrix.

The residual and predicted output error vectors $\bar{\gamma}_{k+1}$ and $\gamma_{k+1}$ are defined as Eq. (6) and Eq. (7).

$$\bar{\gamma}_{k+1} = y_{k+1} - \varphi_{k+1}\hat{\theta}_{k+1} \tag{6}$$

$$\gamma_{k+1} = y_{k+1} - \varphi_{k+1}\hat{\theta}_k \tag{7}$$

Predicted output error covariance matrix is denoted by $V_{k+1} = E[\gamma_{k+1}\gamma_{k+1}^T]$, and when $\hat{\theta}_{k+1}$ reaches to $\theta_{k+1}$. It would be

$$E[\bar{\gamma}_{k+1}\bar{\gamma}_{k+1}^T] = E[\varepsilon_{k+1}\varepsilon_{k+1}^T] = \sigma_{k+1}^2 \tag{8}$$

The adaptive tracking condition has been obtained as Eq. (9) follows in [5].

$$\begin{aligned} &V_{k+1} - [I + \varphi_{k+1}(\Lambda_{k+1} P_k \Lambda_{k+1}^T)\varphi_{k+1}^T]\sigma_{k+1}^2 \\ &[I + \varphi_{k+1}(\Lambda_{k+1} P_k \Lambda_{k+1}^T)\varphi_{k+1}^T]^T = 0 \end{aligned} \tag{9}$$

$\varphi_{k+1}$ is measured and $P_k, V_{k+1}$ and $\sigma_{k+1}^2$ are estimated. Eq. (9) is nonlinear, and it is difficult to find a solution. The objective function in Eq. (10) is used to find the optimal solution [5].

$$J[\hat{\theta}_{k+1}(\Lambda_{k+1})] = \sum_{j=1}^{n} \left| \frac{\hat{\theta}_j(k+1) - \hat{\theta}_j(k)}{\hat{\theta}_j(k)} \right| \tag{10}$$

Objective function (10) is the adding of the variation of parameters from $\hat{\theta}(k)$ to $\hat{\theta}(k+1)$. Finding the optimal solution $\Lambda_{k+1}$ is a constrained optimization problem with the objective function (10) subject to the constraint of the Eq. (9) norm,

$$\begin{aligned} &\left\| V_{k+1} - [I + \varphi_{k+1}(\Lambda_{k+1} P_k \Lambda_{k+1}^T)\varphi_{k+1}^T]\sigma_{k+1}^2 \right. \\ &\left. [I + \varphi_{k+1}(\Lambda_{k+1} P_k \Lambda_{k+1}^T)\varphi_{k+1}^T]^T \right\| \le \delta \end{aligned} \tag{11}$$

The function "fmincon" in MATLAB is used to find an optimal solution for the adaptive factor matrix $\Lambda_{k+1}$. The initial value for $\Lambda_{k+1}$ is supposed to be $\bar{\lambda}_{k+1}^{1/2} I$, where $\bar{\lambda}_{k+1}/[(V_{k+1}/\mu\sigma_{k+1}^2)^{1/2} - 1]/(\varphi_{k+1} P_k \varphi_{k+1}^T)$ . If the calculated $\bar{\lambda}_{k+1}$ is smaller than one, it is set to be one. In this case all the parameters are constant at $t_{k+1}$.

## 4. Numerical Simulations

A nonlinear hysteretic structural system with one degree of freedom subject to earthquake acceleration $\ddot{x}_0(t)$ is considered.

$$m\ddot{x}(t) + r(\dot{x}, x) = -m\ddot{x}_0(t) \tag{12}$$

where $x$ is the relative displacement and $r(\dot{x}, x) = F_c(\dot{x}) + F_s(x)$ is the total restoring force in which $F_c(\dot{x}) = c\dot{x}$. Bouc-Wen model is used to describe $r(\dot{x}, x)$ [12]:

$$\dot{r} = c\ddot{x} + k\dot{x} - \beta\left|\dot{x}\right|\left|r\right|^{n-1} r - \gamma\left|\dot{x}\right|\left|r\right|^{n} \tag{13}$$

where $c$ is the damping coefficient, $k$ is the equivalent stiffness, and $\beta$, $n$ and $\gamma$ are hysteresis parameters. Values of the parameters used in the simulation are shown in Table 1 ([3,5]). The El-Centro earthquake with a $5g$ peak ground acceleration is used. For measured quantities the assumed sampling time is 1$KHz$.

To identify the parameters, Eq. (12) and Eq. (13) must be discrete and be converted to observation Eq. (3). The incremental component of restoring force $r_k$ based on a 3rd order corrector method could be expressed as Eq. (14).

$$r_k - r_{k-1} = (\Delta t / 12)(5\dot{r}_k + 8\dot{r}_{k-1} - \dot{r}_{k-2}) \tag{14}$$

The unknown parameter vector is defined as a 4-vector $\theta_k = [c, k, \beta, \gamma]^T$. The me asured vector $y_k$ can be computed from the measured data and is defined as Eq. (15).

$$\begin{aligned} &y_k = (12 / \Delta t)(r_k - r_{k-1}) = \\ &(-12m / \Delta t)(\ddot{x}_k - \ddot{x}_{k-1} + \ddot{x}_{0,k} - \ddot{x}_{0,k-1}) \end{aligned} \tag{15}$$

The data matrix $\varphi_k = [\varphi_{k,1}\ \varphi_{k,2}\ \varphi_{k,3}\ \varphi_{k,4}]$ could be obtained as follows:

$$\varphi_{k,1} = 5\ddot{x}_k + 8\ddot{x}_{k-1} - \ddot{x}_{k-2}$$
$$\varphi_{k,2} = 5\dot{x}_k + 8\dot{x}_{k-1} - \dot{x}_{k-2}$$
$$\varphi_{k,3} = 5 \mid \dot{x}_k \mid\mid m(\ddot{x}_k + \ddot{x}_{0,k}) \mid^{\alpha-1} m(\ddot{x}_k + \ddot{x}_{0,k})$$
$$+8 \mid \dot{x}_{k-1} \mid\mid m(\ddot{x}_{k-1} + \ddot{x}_{0,k-1}) \mid^{\alpha-1} m(\ddot{x}_{k-1} + \ddot{x}_{0,k-1})$$
$$- \mid \dot{x}_{k-2} \mid\mid m(\ddot{x}_{k-2} + \ddot{x}_{0,k-2}) \mid^{\alpha-1} m(\ddot{x}_{k-2} + \ddot{x}_{0,k-2})$$
$$\varphi_{k,4} = 5\dot{x}_k \mid m(\ddot{x}_k + \ddot{x}_{0,k}) \mid^{\alpha} -8\dot{x}_{k-1} \mid m(\ddot{x}_{k-1} + \ddot{x}_{0,k-1}) \mid^{\alpha}$$
$$+\dot{x}_{k-2} \mid m(\ddot{x}_{k-2} + \ddot{x}_{0,k-2}) \mid^{\alpha} \quad (16)$$

Data matrix includes structural responses $\ddot{x}_k$, $\dot{x}_k$ and also the earthquake acceleration $\ddot{x}_{0,k}$. Seismic acceleration $\ddot{x}_{0,k}$ and acceleration response $\ddot{x}_k$ are measured using accelerometers. Velocity response $\dot{x}_k$ can be calculated using numerical integration of $\ddot{x}_k$. The hysteretic structural system parameters are shown in Table 1. The system stiffness $k$ decreases suddenly at the moment of $t=15s$ from *24.2kN/m* to *20kN/m* due to failure. Initial values were set to $c_0 = 0.1kNs/m$, $k_0 = 10kN/m$, $\beta_0 = 0$, $\gamma_0 = 0$, $P_0 = 100I$.

The identification results of a single degree of freedom structural system (SDOF) parameters using LSE algorithm, fixed and variable forgetting factor and the adaptive algorithm based on LSE and optimization are shown in Fig. 1 to Fig. 4. The time part $t<2$ is not used for parameter identification, because the earthquake and its response are too small in this segment.

As shown in Fig. 1 the LSE algorithm can be used to identify fixed parameters, but this algorithm can not correctly identify the time-varying parameters. The constant forgetting factor which is used and shown in Fig. 2 modifies the variable parameter identification to a large extent, but other parameters are not well identified.

LSE with variable forgetting factor algorithm improves the results of the algorithm with a constant forgetting factor in Fig. 3. However, the identification of these two methods works regardless of which parameter changes and the oscillations seen in Fig. 2 and Fig. 3.

Parameter identification in adaptive algorithm based on LSE and optimized adaptive factor matrix is performed by setting the optimal coefficients and the identification error is very little in Fig. 4. Also the exact and identified hysteresis cycle with taking the stiffness variation in $t=15s$ are shown in Fig. .

Table 1. Hysteretic system parameters

| Parameter | Value |
|---|---|
| $m, c, k$ $\beta, \gamma, n$ | *125.53kg, 0.07kNs/ m, 24.2kN/ m 2, 1, 2* |

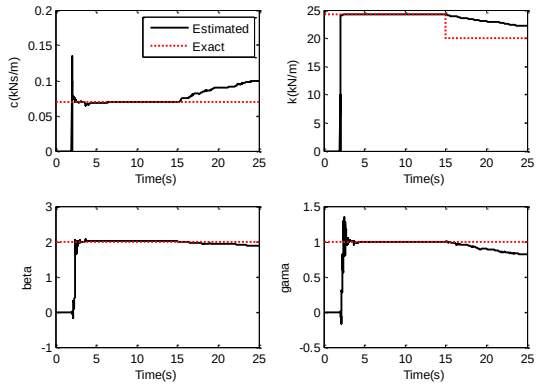



Fig. 1. Parameters $c, k, \beta, \gamma$ identified using least-squares estimation algorithm.

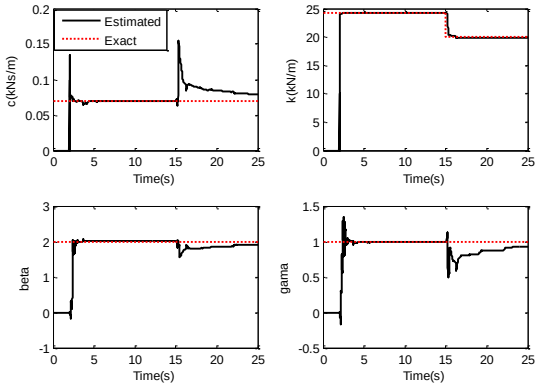



Fig. 2. Parameters $c, k, \beta, \gamma$ identified using least-squares estimation algorithm with constant forgetting factor.

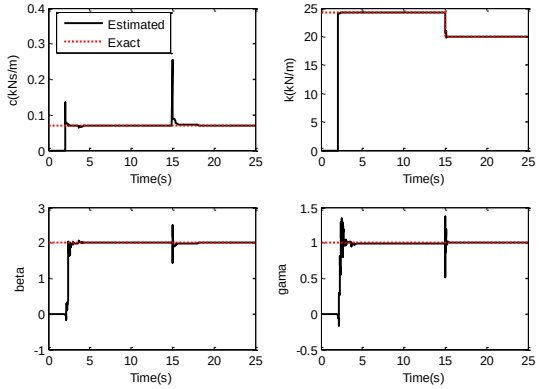



Fig. 3. Parameters $c, k, \beta, \gamma$ identified using least-squares estimation algorithm with variable forgetting factor.

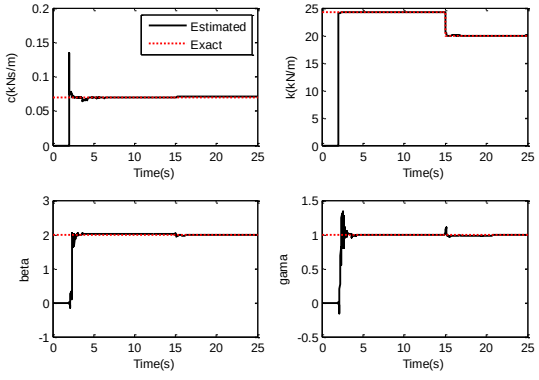



Fig. 4. Parameters $c, k, \beta, \gamma$ identified using adaptive least-squares estimation algorithm with optimization.

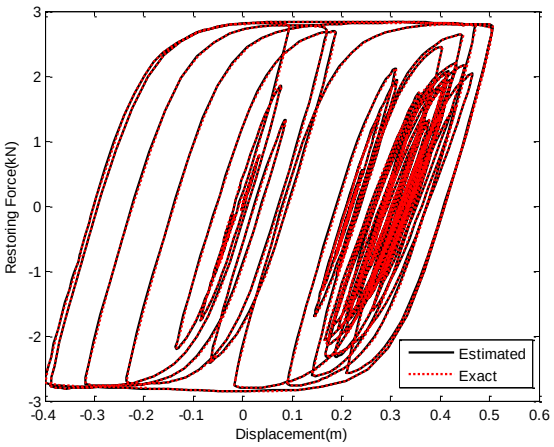



Fig. 5. Estimated and exact hysteresis cycles for a system with one degree of freedom with a stiffness loss.

## 5. Conclusions

Different identification algorithms based on least-squares estimation are used for identification of variable parameters of a hysteretic structural system. Efficiency of the adaptive coefficient matrix method is shown using the results of numerical simulations and compared with the results of ordinary least-squares estimation algorithms, fixed and variable forgetting factor algorithm. Results indicate that the adaptive coefficient matrix algorithm compared with other methods have better performance especially at the fault moment modeled as the immediate stiffness parameter change. Evaluation of the method to measurement noise, application to more complex structures and improvement of the optimization method are some future research topics.

## References

[1] S.W. Doebling, C.R. Farrar, M.B. Prime, A summary review of vibration-based damage identification methods, Shock and Vibration Digest, 30 (1998) 91-105.
[2] Q.W. Yang, Review of vibration-based structural damage identification methods, Zhendong yu Chongji/Journal of Vibration and Shock, 26 (2007) 86-91.
[3] J.W. Lin, R. Betti, A.W. Smyth, R.W. Longman, On-line identification of non-linear hysteretic structural systems using a variable trace approach, Earthquake Engineering and Structural Dynamics, 30 (2001) 1279-1303.
[4] H. Luş, R. Betti, R.W. Longman, Identification of linear structural systems using earthquake-induced vibration data, Earthquake Engineering and Structural Dynamics, 28 (1999) 1449-1467.
[5] J.N. Yang, S. Lin, On-line identification of non-linear hysteretic structures using an adaptive tracking technique, International Journal of Non-Linear Mechanics, 39 (2004) 1481-1491.
[6] R. Ghanem, M. Shinozuka, Structural-system identification. I: Theory, Journal of Engineering Mechanics, 121 (1995) 255-264.
[7] M. Shinozuka, R. Ghanem, Structural system identification. II: Experimental verification, Journal of Engineering Mechanics, 121 (1995) 265-273.
[8] T. Sato, K. Qi, Adaptive H∞ filter: its application to structural identification, Journal of Engineering Mechanics, 124 (1998) 1233-1240.
[9] C.-C. Chang, Y. Shi, Identification of time-varying hysteretic structures using wavelet multiresolution analysis, International Journal of Non-Linear Mechanics, 45 (2010) 21-34.
[10] M. Peimani, M.J. Yazdanpanah, N. Khaji, Parametric identification of hysteretic systems using least-squares estimation based methods, in: 21st Iranian Conference on Electrical Engineering Ferdowsi Univ. of Mashhad, 2013.
[11] F. Ikhouane, J. Rodellar, Systems with Hysteresis: Analysis, Identification and Control Using the Bouc-Wen Model, John Wiley & Sons Ltd, UK, 2007.
[12] Y.-K. Wen, Method for random vibration of hysteretic systems, ASCE J Eng Mech Div, 102 (1976) 249-263.
[13] F. Qian, S. Leung, Y. Zhu, W. Wong, D. Pao, W. Lau, Damped sinusoidal signals parameter estimation in frequency domain, Signal Processing, 92 (2012) 381-391.
[14] H.K. Kim, J.H. Yoon, Y. Li, Asymptotic properties of least squares estimation with fuzzy observations, Information Sciences, 178 (2008) 439-451.
[15] F.K.W. Chan, H.C. So, W.H. Lau, C.F. Chan, Structured total least squares approach for efficient frequency estimation, Signal Processing, 91 (2011) 1043-1047.
[16] G.C. Goodwin, K.S. Sin, Adaptive Filtering, Prediction and Control, Prentice- Hall, Englewood Cliffs, NJ, 1984.

**Mansour Peimani** was born in Tabriz, Iran on January 17, 1982. He received the B.Sc. and M.Sc. degrees in Electrical Engineering from University of Tabriz, Iran in 2004 and 2006. He is currently Ph.D. candidate of Electrical Engineering in Science and Research Branch, Islamic Azad University, Tehran, Iran. His research interests are System Identification, Nonlinear Control Systems and Fault Diagnosis. From 2007 till present, he is an academic staff member in Department of Electrical Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran.

**Mohammad Javad Yazdanpanah** received his B.Sc., M.Sc., and Ph.D. degrees all in Electrical Engineering from Isfahan University of Technology, Isfahan, Iran in 1986, Tehran University, Tehran, Iran, in 1988, and Concordia University, Montreal, Quebec, Canada in 1997, respectively. His Ph.D. thesis entitled “Control of flexible-link manipulators using nonlinear H∞ techniques” was ranked outstanding.
From 1986 to 1992, he worked with the Engineering Research Center, Tehran, Iran, as an R&D engineer and culminating as the chairman of the System Design Division.
In 1998, he joined the School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran, where he is now a Professor, director of the Control & Intelligent Processing Center of Excellence (CIPCE), and director of the Advanced Control Systems Laboratory (ACSL).

Dr. Yazdanpanah's research interests are in the areas of analysis and design of nonlinear/ robust/ optimal/ adaptive control systems, robotics, control on networks, and theoretical and practical aspects of neural networks.

**Naser Khaji** is Associate Professor of Earthquake Engineering at Tarbiat Modares University, Tehran, Iran, since 2002 where he has been teaching Advanced Engineering Mathematics, Finite Element Method, Seismic Hazard Analysis, and Boundary Element Method. He earned his B.Sc. degree in Civil Engineering from Tehran University, Iran, in 1995, his M.Sc. degree in Hydraulic Structures from Tarbiat Modares University, Iran, in 1998, and his Ph.D. degree in Earthquake Engineering from the University of Tokyo, Japan, in 2001. Dr. Khaji has published 40 papers in peer-reviewed national and international journals, and numerous conference articles. Dr. Khaji's research interests include: Computational Mechanics, Fluid-Soil-Structure Interaction, Engineering Seismology, and Health Monitoring of Structures.

# Digital Video Stabilization System by Adaptive Fuzzy Kalman Filtering

Mohammad Javad Tanakian*
Communication, M.Sc, Faculty of Electrical & Computer Engineering, University of Sistan and Baluchestan
j.tanakian@gmail.com
Mehdi Rezaei
Signal Processing, Assistant Professor, Faculty of Electrical & Computer Engineering, University of Sistan and Baluchestan
mehdi.rezaei@ece.usb.ac.ir
Farahnaz Mohanna
Image Processing, Assistant Professor, Faculty of Electrical & Computer Engineering, University of Sistan and Baluchestan
f_mohanna@ece.usb.ac.ir

Received: 24/Feb/2012 Accepted: 11/Dec/2013

**Abstract**
Digital video stabilization (DVS) allows acquiring video sequences without disturbing jerkiness, removing unwanted camera movements. A good DVS should remove the unwanted camera movements while maintains the intentional camera movements. In this article, we propose a novel DVS algorithm that compensates the camera jitters applying an adaptive fuzzy filter on the global motion of video frames. The adaptive fuzzy filter is a Kalman filter which is tuned by a fuzzy system adaptively to the camera motion characteristics. The fuzzy system is also tuned during operation according to the amount of camera jitters. The fuzzy system uses two inputs which are quantitative representations of the unwanted and the intentional camera movements. Since motion estimation is a computation intensive operation, the global motion of video frames is estimated based on the block motion vectors which resulted by video encoder during motion estimation operation. Furthermore, the proposed method also utilizes an adaptive criterion for filtering and validation of motion vectors. Experimental results indicate a good performance for the proposed algorithm.

**Keywords:** Adaptive, Digital Video Stabilization, Fuzzy Filter, Kalman Filter, Motion Estimation, Motion Vector, Video Coding.

## 1. Introduction

Digital video stabilization (DVS) techniques have been studied for decades to improve visual quality of image sequences captured by compact and light weight digital video cameras. When such cameras are hand held or mounted on unstable platforms, the captured video generally looks shaky because of undesired camera motions. Unwanted video vibrations would lead to degraded view experience and also greatly affect the performances of applications such as video encoding [1-4] and video surveillance [5,6]. With recent advances in

Wireless technology, video stabilization systems are also considered for integration into wireless video communication equipment for the stabilization of acquired sequences before transmission, not only to improve visual quality but also to increase the compression performance [1]. Solutions to the stabilization problem involve either hardware or software to compensate the unwanted camera motion. The hardware-based stabilizers are generally expensive and lack the kind of compactness that is crucial for today's consumer electronic devices [7, 8]. On the contrary, a DVS system that is implemented by software can easily be miniaturized and updated. Consequently, DVS system is suitable for portable digital devices, such as digital camera and mobile phone.

In general, a DVS system consists of two principal units including motion estimation (ME) and motion correction (MC) units. The ME unit estimates a global motion vector (GMV) between every two consecutive frames of the video sequence. Using the GMVs, the MC unit then generates smoothing motion vectors (SMVs) needed to compensate the frame jitters and warp the frames to create a more visual stable image sequence.

According to the motion models being considered, the already proposed global ME techniques for DVS system can roughly be divided into two categories:

(1) Two- dimensional stabilization techniques which deal with translational jitter only [9-20] and

(2) multi-dimensional stabilization techniques which aim at stabilizing more complicated fluctuations in addition to translation [21-25]. Most of the existing algorithms fall into the first category because the translation is the most commonly encountered motion and the complexity of estimating translation parameters is relatively low for real-time stabilization. In the second category, the majority of algorithms [21,22,24] considered a three-dimensional or perspective motion model, while a few algorithms considered affine motion

* Corresponding Author

model [23] or sensors attached to the camera to provide absolute 3D orientation [25].

Regarding to the two-dimensional ME task of DVS systems, most previous approaches attempt to reduce the computational cost by using fast ME algorithms, e.g. gray-coded bit-plane matching [9], two-bit transform [10], multiplication-free one-bit transform [11], Laplacian two-bit transform [12], and binary image matching of color weight [13]. In another approach, the global ME is limited to small, pre-defined regions [16,17]. Such approaches consider DVS and video encoding separately and attempt to trade the accuracy of motion vectors (MVs) for the computational efficiency; nevertheless they improve the computational efficiency at the expense of degradation in the accuracy in ME and thereafter in MC tasks.

Since both the video encoder and the digital stabilizer of a digital video camera use a ME unit, we can integrate digital stabilizer with video encoder [2,4,26] by making the two modules of a digital video camera share a common local motion vectors (LMVs) estimation process, as shown in Fig. 1. ME can take up 90% of the total computation of a digital stabilizer and 50% -70% of a video codec. Combined together, the operation required for ME can consume more than 70% of the total computation [4]. The ME task in video encoders usually is implemented on frame blocks by a block matching process to estimate a MV for each block (BMV).

The ME unit plays an important role in DVS system and its estimation accuracy is a decisive factor for the overall stabilization performance of the system. In video frames with smooth or complex texture regions, the estimated BMVs may not be in coincidence with the real motion of the blocks. Although such LMVs are applicable to the local motion compensation task which is executed in the encoder, they cannot be used for the global motion compensation which is executed by the DVS. These LMVs include some noises that degrade the global ME task. In order to remove the noisy LMVs in these regions some algorithms are proposed in [27-30]. The valid BMVs as LMVs are used for the global ME and MC compensation in next steps.

After global ME, the next essential task of a DVS system is MC in which the unwanted camera jitters are separated and removed from the intentional camera movement. Among the various MC algorithms proposed in the literature, smoothing of the GMV by low-pass filtering is the most popular. For instance, an MV integration method is used in [9,31] which utilizes a first-order infinite impulse response (IIR) low-pass filter to integrate

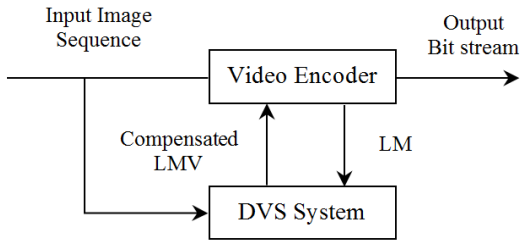



Fig.1 Integration scheme of the video stabilizer and the video encoder.

Differential motion and to smoothen the global movement trajectory. A frame position smoothing (FPS) algorithm, based on smoothing absolute frame positions that achieve successful stabilization performance with retained smooth camera movements, is utilized for MC in [17,32-41]. Off-line discrete Fourier transform (DFT) domain filtering is proposed for FPS-based stabilization in [32].

Kalman filter and fuzzy systems have widely been used in DVS applications [33-41]. A real-time FPS-based stabilizer using Kalman filtering of absolute frame positions has been proposed in [17,33]. In the presented algorithm in [34] two Kalman filters are operated in parallel, one of which is used as a reference filter with a constant high process noise variance and another one is used as stabilization filter in which a fuzzy system lets the process noise variance to be adjusted. In this case, the process noise variance of the stabilization filter is adaptively changed by the fuzzy system according to the residual between the stabilization and reference filter output. Presented DVS in [35] utilizes an fuzzy system in which membership functions (MFs) are optimized to motion dynamics [35]. A membership selective fuzzy stabilization, in which the stabilization system selects between a pre-determined set of MFs according to instantaneous motion characteristics is proposed in [36]. A MF adaptive fuzzy filter based on smoothing of absolute frame position for video stabilization is presented in [37]. In this method initially a short mean filter is applied to raw absolute frame displacement as pre-process, to reduce the dynamic range of the fuzzy system input. Fuzzy stabilization is then achieved through fuzzy correction mapping. In this method output MFs of the fuzzy system are continuously adapted so as to constitute a MF adaptive fuzzy filtering process. It is shown in [38] that the performance of the Kalman filter can be improved by a fuzzy system that improves the Kalman filter output. In presented algorithm in [33,39], the process noise variance in Kalman filter is adjusted adaptively according to the camera motion characteristics.

Almost all fuzzy systems utilized by presented algorithms [34-38] have a similar structure with two inputs as fuzzy inputs and a little difference in fuzzy MFs. In these algorithms, the first input is the difference between the absolute frame displacement and the a priori estimate of the stabilized frame position that achieve by a Kalman filter or another predictor. The second input indicates the change of first input over the last two frames.

Regarding to the MC task of DVS systems, almost all published algorithms try to smoothen the global movement trajectory by a kind of low-pass filtering. An important drawback of the low-pass filtering is that smoothened movement trajectory is delayed with respect to the desired camera displacements. A stricter filtering provides more stabilization at the expense of more trajectory delay and vice versa. More trajectory delay means losing more image content after stabilization.

In this article, we propose a DVS algorithm with new features in ME and MC units. The ME unit estimates a

GMV based on the BMVs which are estimated by the video encoder. Therefore, the computational complexity of the DVS is very low and the accurate motion information is used without extra computation cost. Moreover, in order to improve the accuracy of GMV estimation task an adaptive thresholding algorithm is used to remove the noisy invalid LMVs. The MC unit of the proposed DVS system is an adaptive fuzzy filter that applied on the global motion of video frames to smooth the camera movement trajectory adaptively. The adaptive fuzzy filter is a Kalman filter which is tuned by a fuzzy system adaptively to the characteristics of unwanted and intentional camera motions. The fuzzy system itself is also tuned during operation according to the amount of camera jitters. The fuzzy system uses two inputs which are quantitative representations of the unwanted and the intentional camera movements. Experimental results show a good performance for the proposed DVS algorithm. Also in order to have a generic comparison between our method and some relevant MC algorithms proposed in the literature, Table1 is presented. A good MC unit should remove the unwanted camera motion while tracks the intentional motion without any delay. For this purpose, it should discriminate the unwanted and intentional camera motions while adjust the smoothing filter adaptively according to the amount of unwanted and intentional camera motions. The studied published MC algorithms lack some of these features. For example in the Table1, it is shown that the algorithms presented in [27,39,40] suffer from the lack of discrimination of unwanted and intentional camera motions. Moreover, the proposed adaptive algorithm in [27] suffers from a continuous and well adaptation. They use an adaptive filter with a smoothing factor that is switched between only two values and therefore it leads to undesirable jumps in frame position while our proposed MC algorithm discriminate between the unwanted and intentional camera motions and also uses a fuzzy system to improve the quality of adaptive filter.

The remainder of this article is organized as follows. The details of the proposed video stabilization algorithm are described in Section 2. Some experimental results are presented in Section 3, and the article is concluded in Section 4.

Table 1: Comparison results for some MC algorithms proposed in the literature and our Method

| Criteria | Adaptive Filter | Kalman Filter | Fuzzy System | Discriminating Unwanted and Intentional Motions |
|---|---|---|---|---|
| Adaptive IIR Filter [27] | Y | | | |
| Adaptive Kalman Filter [33] | Y | Y | | |
| Fuzzy Adaptive Kalman Filter [34] | Y | Y | Y | |
| MF Adaptive Fuzzy Filter [37] | Y | | Y | |
| Recursive Fuzzy System [38] | | Y | Y | |
| Adaptive Kalman Filtering [39] | Y | Y | | |
| Adaptive Kalman Filtering [40] | Y | Y | | |
| Proposed Method | Y | Y | Y | Y |

## 2. The Proposed Method

A flowchart of the proposed DVS system is depicted in Fig.2. The details of the proposed system are described in the sequel.

### 2.1 Block-Based ME

The block-based ME is used to generate the LMVs. Since the ME is done by the video encoder, the computational complexity of the DVS is very low and the accurate motion information is used without extra computation cost. In this article, to test the proposed DVS system independent of the encoder, a full search ME algorithm with full-pixel resolution is taken for 8×8 blocks over a search range of 33×33 pixel to achieve the BMVs.

The ME algorithm works as follows. First, the current frame is divided into a number of $N \times N$ blocks and an MV for each block is computed. The resulting MV points to the most correlated reference block in the previous frame within the search area. To measure the goodness of each candidate MV (x,y), the mean absolute difference (MAD) measure is used as

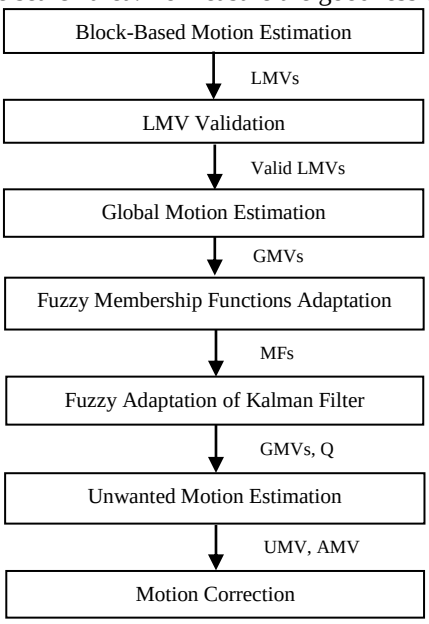



Fig. 2 Flowchart of the proposed DVS system.

$$MAD(i,j) = \frac{1}{N^2}\sum_{K=0}^{N-1}\sum_{L=0}^{N-1}|C(x+k,y+k) - R(x+i+k,y+j+l)| \quad (1)$$

where C(x+k, y+L) and R(x+i+k, y+j+L) denote the block pixels in the target frame, and the displaced block pixels in the reference frame, respectively. The candidate MV (i,j) with the smallest MAD is chosen as the MV of the current block according to:

$$V = \arg_{(i,j)} \min \text{MAD}\,(i,j), -p \le i,j \le p-1 \quad (2)$$

Where *p* defines the motion search range.

## 2.2 LMV Validation

The ME unit plays an important role in DVS system and its estimation accuracy is a decisive factor for the overall performance of stabilization system. Block ME process typically computes some wrong MVs which are not in coincidence to the real motion direction of the blocks. Although, such MVs can be useful for the motion compensation in encoder, they include noise and should not be used for the global motion compensation and video stabilization operations. These LMVs include some noises that degrade the global ME task. The noisy MVs are mostly obtained from two types of regions including: very smooth regions with lack of features and very complex uneven regions [27-30]. Inspiring from the algorithm presented in [27], two qualifying tests, namely "Smoothness Test" and "Complexity Test", are used to detect and remove the noisy MVs by an adaptive thresholding method as follows. The valid BMVs as LMVs are used for the global ME and MC in next steps.

### 2.2.1 Smoothness Test

The noisy MVs corresponding to the smooth regions such as sky image are detected by thresholding of the average of MAD as:

$$\text{MAD}^n_{\text{avg}} < \text{th}_1, \quad (3)$$

where $\text{MAD}^n_{\text{avg}}$ denotes the average of calculated MADs within the search area, during ME of $n^{th}$ block. $th_1$ is also defined as

$$\text{th}_1 = \text{MAD}^n_{\text{min}} + T_1 \times \text{Mean}(\text{MAD}^n_{\text{avg}}) \quad (4)$$

Where $\text{MAD}^n_{\text{min}}$ and $\text{MAD}^n_{\text{avg}}$ denote the minimum and the average values of computed MADs, respectively, during ME of $n^{th}$ block within the search area. $T_1$ is an experimentally defined constant coefficient about 0.45 and $\text{Mean}(\text{MAD}^n_{\text{avg}})$ denotes the average of $\text{MAD}^n_{\text{avg}}$, over all blocks of the frame. In fact the threshold $th_1$ includes a global average value over the frame plus a margin.

### 2.2.2 Complexity Test

The noisy MVs corresponding to the complex texture regions are identified by another thresholding as:

$$\text{MAD}^n_{\text{min}} > \text{th}_2, \quad (5)$$

Where threshold $th_2$ is defined adaptively as:

$$\text{th}_2 = T_2 \times \text{Max}(\text{MAD}^n_{\text{min}}), \quad (6)$$

Where T2 is an experimentally defined constant coefficient about 0.45, and $\text{Max}(\text{MAD}^n_{\text{min}})$ denotes the maximum value of $\text{MAD}^n_{\text{min}}$, over all blocks of the frame. According to the equations above, the $\text{MAD}^n_{\text{min}}$ is compared against a portion of its global maximum over a frame.

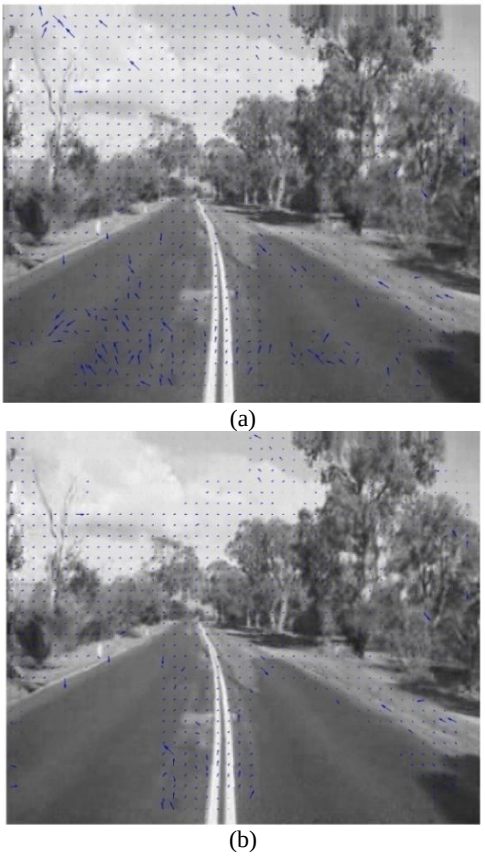
It is notable that MAD is computed during ME by encoder. Therefore, the smoothness test and complexity test have no additional computational complexity cost for the proposed DVS system.

Provided results on different video contents show that, using fixed thresholds for different video contents may cause a remarkable amount of invalid noisy LMVs remain or a notable amount of valid LMVs be removed. To solve this problem, the values of thresholds $th_1$ and $th_2$ are adjusted adaptively based on the video content for each frame. Note, if ME is executed by a fast search algorithm rather than full-search algorithm at the encoder, the MADs calculated during ME are used for adaptation of thresholds $th_1$ and $th_2$.

Original LMVs and validated LMVs for a sample frame are presented in Fig.3. This figure shows that many noisy LMVs have been removed by the LMV validation process.

## 2.3 Global ME

The global ME unit produces a unique GMV for each video frame, which represents the camera movement during the time interval of two frames. Since the LMVs obtained from the image background tend to be very similar in both magnitude and direction, we used a clustering process to classify the motion field into clusters corresponding to the background and foreground objects.



(a) (b)

Fig. 3 Example of noisy LMV removal on a frame of the avenue sequence. (a) Original MVs, (b) Valid MVs.

The global motion induced by camera movement is determined by a clustering process that consists of the following steps.

**Step 1)** Construct the histogram H of the valid LMVs. The value of H(x,y) is incremented by one each time the LMV(*x*,*y*) is encountered.

**Step 2)** As long as the scene is not dominated by moving objects, the cluster corresponding to background blocks has the maximum votes in the clustering process. The position (*x*,*y*) of the largest cluster or histogram bin is considered as the GMV.

As an example, Fig. 4 shows the largest histogram bin at coordinates (5,12), yields the GMV.

## 2.4 Unwanted ME and Correction

An estimated GMV may consist of two major components: an intentional motion component (e.g., corresponding to camera panning) and unintentional motion component (e.g., corresponding to handshake). A good MC algorithm should only remove the unwanted motion while maintain and track the intentional motion. Assuming that the unwanted motion is corresponding to the high-frequency components, the proposed algorithm uses a low-pass filter to remove the unwanted motion component. An SMV is resulted by a Kalman filtering on the GMVs that resembles the intentional camera movement. The Kalman filter provides an estimation to the state of a discrete-time process defined as a linear dynamical system as

$$x(t+1) = F \times x(t) + w(t). \quad (7)$$

Where *F* and *w(t)* represent state transition matrix and the process noise, respectively. The Kalman filter operates

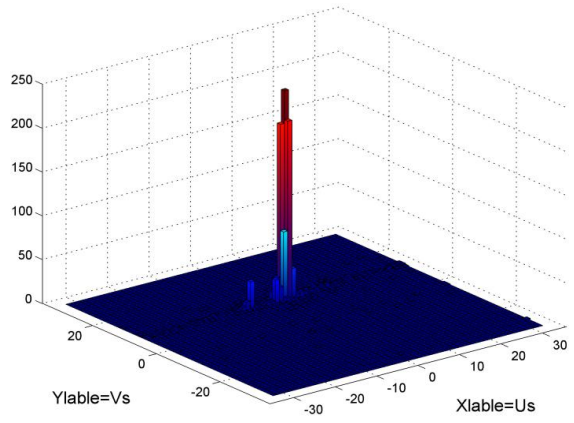



Fig. 4 Clusters of Motion field

using observations defined by the observation system as

$$y(t) = H \times x(t) + v(t). \quad (8)$$

Where *H* and *v(t)* show measurement matrix and noise, respectively. Process and measurement noise are assumed to be independent of each other, white, and with normal probability distributions: $w \sim N(0, Q)$ and $v \sim N(0, R)$.

The Q (process noise variance) in Kalman filter has a direct effect on the operation of stabilizer. A relative small variance value provides expanded stabilization with reduced adaptability to changes in intentional motion dynamics, it means Kalman acts as a strong filter and provides expanded stabilization. While a relative large variance value enables close tracking of intentional camera movements but at the cost of slightly reduced stabilization capabilities, it means Kalman acts as a weak filter and quickly adjusts to changes in intentional motion dynamics. So, we can have an adaptive Kalman filter if the process noise variance of Kalman filter changes automatically during filtering operation. A fixed value of Q not effectively leads to good stabilized image sequences [13]. To avoid the lag of intentional movement and to smooth the unwanted camera motion efficiently, the following fuzzy adaptation mechanism of Q is proposed.

### 2.4.1 Fuzzy Adaptation of Kalman Filter

The Kalman filter is implemented on the vertical and horizontal components of the GMVs separately. The process noise variance of Kalman filter i.e., Q(n) is adjusted by a fuzzy system continuously for MC of each frame. In facts, two fuzzy systems with a similar structure are used corresponding to the vertical and horizontal motion components. The fuzzy system has two inputs (Input1, Input2) and one output. The fuzzy inputs are defined as:

$$x_1 = \frac{1}{M}\sum_{i=n-M+1}^{n} |GMV_x(i) - GMV_x(i-1)|, \quad (9)$$

$$x_2 = |GMV_x(n) - GMV_x(n-M)|. \quad (10)$$

$$y_1 = \frac{1}{M}\sum_{i=n-M+1}^{n} |GMV_y(i) - GMV_y(i-1)|, \quad (11)$$

$$y_2 = |GMV_y(n) - GMV_y(n-M)|. \quad (12)$$

where $x_1$ and $x_2$ denote the inputs of fuzzy system used for the adaptive filtering of the horizontal motion component and also $y_1$ and $y_2$ are the inputs of fuzzy system used for the adaptive filtering of the vertical motion component. $GMV_x(n)$ and $GMV_y(n)$ indicate the horizontal and vertical components of the GMV of last frame and M+1 is the number of last GMVs used for decision. The fuzzy system inputs, Input1 ($x_1$,$y_1$) and Input2 ($x_2$,$y_2$), are used as quantitative representations of unwanted and intentional camera movements, respectively. The value of Input1 is proportional to the noise amplitude and the value of Input2 is proportional to the intentional camera motion when it has an accelerating movement.

Defining suitable inputs for an adaptive DVS system has a great impact on the performance of system. Only relevant inputs can provide precise discriminating between unwanted and intentional camera motions to be used for the adaptation of Process Noise Variance. Different scenarios for the combination of unwanted and intentional camera motion can be considered. As examples, some scenarios are presented graphically in Fig. 5. In graphs (a) and (b), camera has an intentional accelerating movement plus noise or unwanted motion. The noise amplitude is high in (a) while it can be ignored in (b). Graph (e) is corresponding to a camera movement path while panning in which the camera is moving with a constant velocity without any acceleration and noise. The explanations of all graphs are summarized in Table2.

From the adaptive filtering point of view it is important to measure the amount of noise and the

intentional camera movement velocity and acceleration. A softer Process Noise Variance is needed when the noise amplitude is high to remove the noise. On the other hand the soft Process Noise Variance prevents following of camera path when

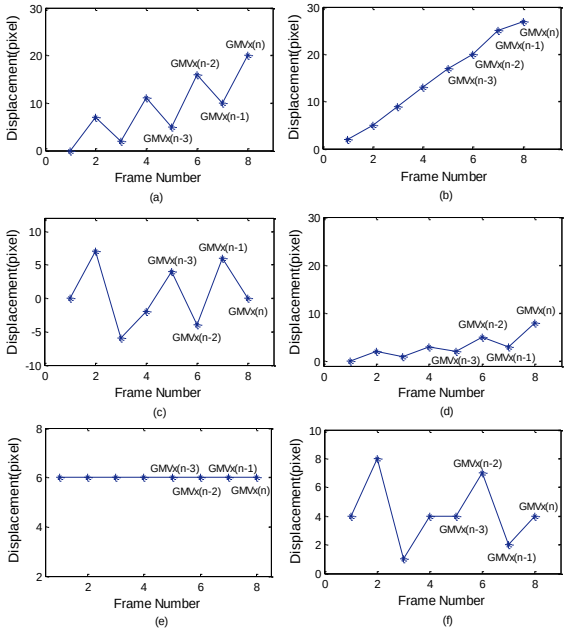



Fig.5 Sample scenarios for combination of unwanted and intentional camera motions: (a) high acceleration with high noise, (b) high acceleration with no noise, (c) high noise with no acceleration, (d) low acceleration with low noise, (e) constant velocity without noise and acceleration, (f) constant velocity with noise.

Table 2: Sample scenarios for combination of unwanted and intentional camera motion

| *Graph* | *Noise* | *Velocity* | *Acceleration* |
|---|---|---|---|
| a | High | High | High |
| b | Low | High | High |
| c | High | Zero | Zero |
| d | Low | Low | Low |
| e | Zero | High | Zero |
| F | High | High | Zero |

it has an intentional high acceleration. Therefore, the Process Noise Variance of Kalman filter should be tuned carefully proportional to the amount of noise and camera movement acceleration. According to this, we defined the fuzzy inputs so that Input1 gives information about the amount of noise and Input2 gives information about the amount of camera movement acceleration. It is notable that amount of camera movement velocity itself does not have any constrain on the filtering so it is not measured and used here. The proposed fuzzy system tunes the Process Noise Variance of the Kalman filter, Q(n), adaptively according to the amount of noise and the camera intentional accelerating movement. In the proposed fuzzy system, trapezoidal and triangular MFs are used for the inputs and the outputs, respectively. The number of MFs has been selected so as to obtain decent performance with as few MFs as possible to maintain low system complexity. The experimentally designed input and output membership functions and also the surface of desired outputs are shown in Fig. 6.

According to experimental results, the performance of used Kalman filter is more sensitive to Q's changes where Q has a small value. Therefore, more MFs of the fuzzy output are concentrated in this operating area. The constructed rule base is containing 30 rules as presented in Table3. The proposed fuzzy system was implemented while the *min* function was used for the *fuzzy implication* and the *max* function used for the *fuzzy aggregation*. Furthermore, the centroid defuzzification method was applied. The output of fuzzy system defines the Q of Kalman filter, i.e., Q (n). for MC of $n^{th}$ video frame.

### 2.4.2 Adaptive Fuzzy MFs

Study on a number of video sequences has shown that the range of fuzzy inputs (Input1, Input2) is very variable on different video contents. Therefore, fixed MFs for the inputs of fuzzy system cannot provide a good stabilization performance over all video contents. In order to have a good performance for the proposed DVS system over different video contents, it is proposed to adjust the MFs of fuzzy inputs adaptively to recently received video frames. The range of MFs for the fuzzy inputs, i.e.,

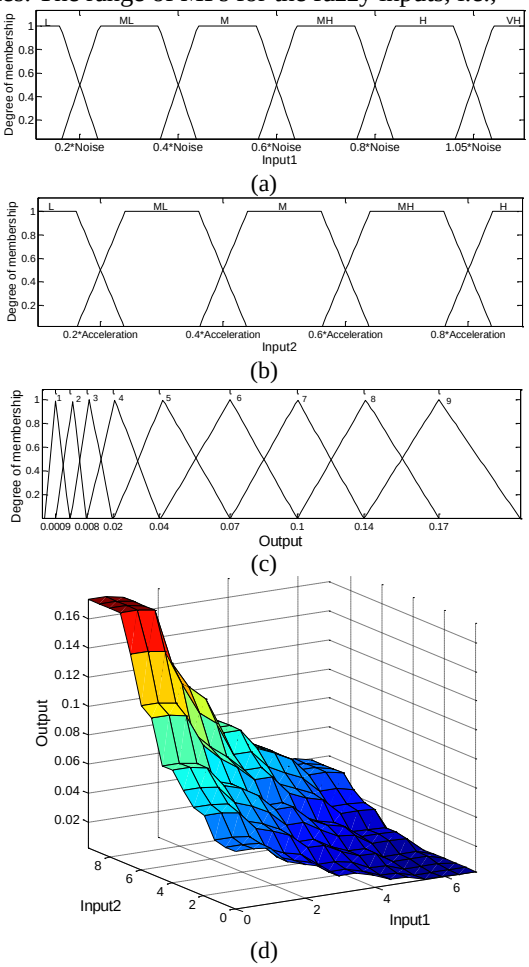



Fig. 6 (a) MFs of fuzzy Input1($x_1$,$y_1$), (b) MFs of fuzzy Input2($x_2$,$y_2$) (c) MFs of fuzzy output, (d) Surface of desired outputs.

(0, Input1(max)) and (0, Input2(max)) are modified adaptively as

Input1(max)= Max of input1over K recent frames, (13)

Input2(max)= Max of input2 over K recent frames, (14)

where Input1 and Input2 are clipped to a range from 1 to 10% of video frame height in term of pixel, and the K corresponds to the number of frames received in last few seconds, e.g., 2 s. This means that the system is adapted to the time-varying noise conditions while the frame size and frame rate are considered.

### 2.4.3 Motion Correction Algorithm

After computing the Q(n) (Process noise variance) by the fuzzy system, the camera motion path constructed by the

Table 3: Central Values of Fuzzy System Output*.

| | | *Input1* | | | | | |
|---|---|---|---|---|---|---|---|
| *Input2* | | L | ML | M | MH | H | VH |
| | L | 0.04 | 0.02 | 0.008 | 0.004 | 0.004 | 0.0009 |
| | ML | 0.07 | 0.04 | 0.02 | 0.008 | 0.004 | 0.0009 |
| | M | 0.1 | 0.07 | 0.04 | 0.02 | 0.008 | 0.0009 |
| | MH | 0.14 | 0.1 | 0.07 | 0.04 | 0.02 | 0.0009 |
| | H | 0.17 | 0.14 | 0.1 | 0.07 | 0.04 | 0.004 |

* L, low; ML, medium low; M, medium; MH, medium high; H, high; VH, very high

GMVs is filtered by the Kalman filter to compute the smoothened motion vectors (SMVs). For the first three frames, a fixed small value for Q(n) is used. After computing SMV, the unwanted motion vector (UMV) is obtained by

$$UMV(n) = GMV(n) - SMV(n). \quad (15)$$

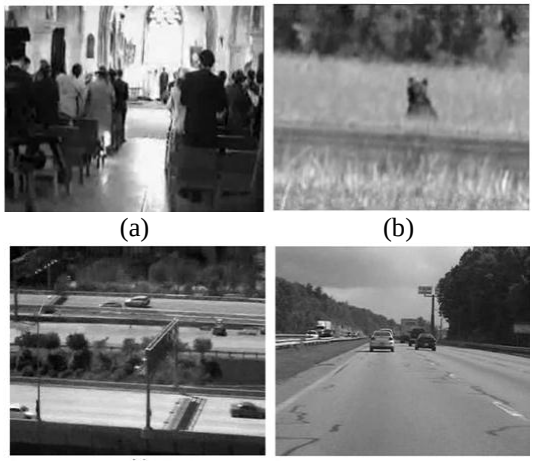
To restore the current frame to its stabilized position, we offset the current frame by the accumulated UMV, AMV, defined by

$$AMV(n) = \sum_{i=m}^{n} UMV(i). \quad (16)$$

Where *m* is the frame number of the last scene cut frame.

## 3. Experimental Results

The performance of the proposed DVS method is evaluated over 15 video sequences covering different types of scenes. Since there is no well-known video sequence in this research field, the algorithm is tested on a number of sequences which are easily available. For example, some used video sequences are available at [42,43]. These sequences have a frame rate of 25 fps and a picture size of 352×288 pixels. Sample frames of used video sequences are shown in Fig. 7. We worked with both gray-scale and color test sequences where in both cases ME is implemented on the luminance component. Good experimental results are obtained with M=3. However, a larger M provides more smoothness at the expense of more tracking delay and vice versa. Performance evaluation of the proposed DVS system are assessed according to the accuracy and computational complexity cost of the GMV estimation and also assessed according to the smoothness of the resultant global motion compared to the original sequence and the gross movement preservation capability.



(a) (b) (c) (d)

Fig. 7 Images taken by a camera (a-c) held by a hand; (d) in a moving vehicle

In this article, we propose a DVS algorithm with new features in ME and MC units. The ME unit estimates a GMV based on the BMVs which are estimated by the video encoder. Therefore, the computational complexity of the DVS is very low and the accurate motion information is used without extra computation cost. Moreover, in order to improve the accuracy of GMV estimation task an adaptive thresholding algorithm is used to remove the noisy invalid LMVs. It is notable that the adaptive thresholding algorithm uses MAD and MAD is computed during ME by encoder. Therefore, the adaptive thresholding algorithm have no additional computational complexity cost for the proposed DVS system. Regarding the mentioned modifications, we believe the ME unit of the proposed DVS outperforms the ME unit of the anchor systems, such as presented algorithms in [38,40], in terms of accuracy and computational cost. Therefore, only the performance of MC unit of proposed DVS system was compared with that of presented algorithms in [38,40] which are the most relevant known competitors. Fuzzy systems and adaptive Kalman filters are utilized for MC unit in [38,40]. The algorithms were applied on several data sets to simulate various scenarios. Since the performance of proposed algorithm in [38] depends on the value of Q parameter (*process noise variance*) of Kalman filter, so it was tuned with two different values of noise variance including Q=0.1 and Q=0.01. Some graphical comparison results are presented in Fig. 8. According to the results, the competitor algorithm proposed in [38] with small Q provides expand stabilization but reduces the capability of close tracking the intentional camera motion. Moreover, with a large Q, it enables close tracking of intentional camera movements

but at the cost of slightly reduced stabilization capabilities. On the other hand,

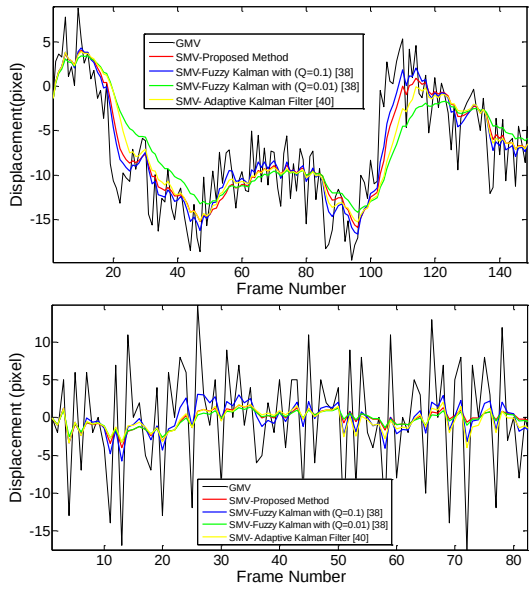



Fig. 8 Comparison results of proposed DVS algorithm with presented algorithm in [38] by two different Kalman filters: Q=0.1, Q=0.01, and presented algorithm in [40].

according to the comparison results, the proposed algorithm in [40] provides expand stabilization but reduces the capability of close tracking the intentional camera motion. Whereas results demonstrate that this article provides good stabilization and close tracking of intentional camera movements at the same time for all cases by adaptive tuning of the noise variance value. Numerical performance assessment of MC unit of a DVS system is a difficult task since the ground truth of unwanted motions as reference is not available. To solve this problem, we produced a synthetic signal as reference. The reference signal includes three parts corresponding to intentional camera movements with positive, negative, and zero value accelerations. Moreover, a sequence of normal random numbers was generated and the random numbers were rounded to integer values to simulate the unwanted camera motion in terms of pixel displacement. Provided sequence was added to the reference signal to obtain a synthetic signal including both the intentional and the unwanted camera motions. Two copies of the synthetic signal ($y_1$,$y_2$) were generated to be processed by MC units of the proposed DVS algorithm and also by the anchor algorithms presented in [38,40]. While the reference signal is known that the performance of compared algorithms can be evaluated numerically by computing a distance measure such as mean square error (MSE) between the reference signal as GMVs of the reference and the processed signals as smoothed GMVs. The MSE measure is computed between the smoothed GMVs and

the GMVs as:

$$\mathrm{MSE_{MV}} = \frac{1}{N}\sum_{n=0}^{N}[\mathrm{GMV_R}(n) - \mathrm{SMV}(n)].^2 \qquad (17)$$

Where $\mathrm{GMV_R}(n)$ and $\mathrm{SMV}(n)$ denote $n^{th}$ GMV of the reference and the smoothed video sequence, respectively. It is noted the performance of proposed algorithm in [38] depends on the value of Q parameter (*process noise variance*) of Kalman filter, so it cannot provide a well daptation. Therefore, the two copies of the synthetic signal were processed by two values of Q (0.005 and 0.6) and graphical simulation results are presented in Fig.9 and Fig. 10, respectively. Moreover, numerical comparison results in term of MSE for the synthetic signals are presented in table 4.

According to the graphical results as shown in Fig.9 and Fig.10, the proposed algorithm in [38], depending on the Q value, performs only good filtering of the unwanted camera motion or only the tracking of intentional camera movement. On the other hand, the proposed algorithm

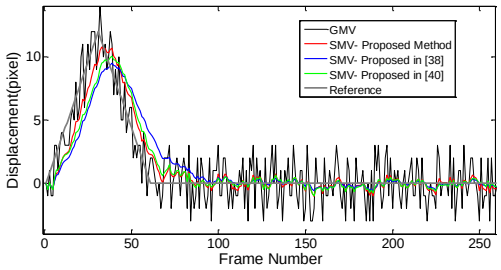



Fig. 9 Synthetic camera movement path $y_1$ processed by MC units of different DVSs

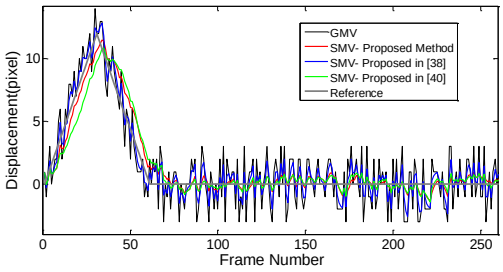



Fig. 10 Synthetic camera movement path $y_2$ processed by MC units of different DVSs

Table 4: MSE resulted by MC units of different DVSs

| ***Motions Signal*** | ***Adaptive Fuzzy Kalman Filter*** *(proposed)* | ***Fuzzy Kalman Filter, [38]*** | ***Adaptive Kalman Filter, [40]*** |
|---|---|---|---|
| **$y_1$** | **0.9941** | **2.3239** | **1.5509** |
| **$y_2$** | **0.6045** | **1.0499** | **1.0579** |

in [40] provides expand stabilization but reduces the capability of close tracking the intentional camera motion, respectively. Whereas the proposed algorithm in this article shows a high performance in both the removing of unwanted motion and the tracking of intentional camera movement. Furthermore, the numerical results presented in Table4 confirm the graphical results shown in Fig. 9 and Fig.10. The least MSE for our proposed algorithm means the best performance. A small-scale subjective quality test also demonstrated that human eyes have better visual perception to the stabilized videos by the proposed DVS system than the original videos in all cases.

## 4. Conclusions

In this article, we proposed a computationally efficient DVS algorithm using motion information obtained from a hybrid block-based video encoder. Since some of the obtained MVs are not valid, an adaptive thresholding was developed to filter out valid MVs and to compute an accurate GMV for each frame. The GMVs are smoothened with a kalman filter that is tuned adaptively to unwanted and intentional camera movements. The filter is adjusted by a fuzzy system with two inputs which quantify the unwanted and intentional camera movements. The proposed method fulfills two apparently conflicting requirements: close follow-up of the intentional camera movement and removal of the unwanted camera motion. In order to improve the stabilization performance, inputs MFs of the fuzzy system are continuously adapted according to the motion properties of a number of recently received video frames. Simulation results show a high performance for the proposed algorithm. With a low degree of computational complexity, the proposed scheme can effectively be used for the mobile video communications as well as for the conventional video coding applications to improve the visual quality of digital video and to provide a higher compression performance.

## References


[1] A. Engelsberg and G. Schmidt,"A comparative review of digital image stabilizing algorithms for mobile video communications", IEEE Trans. Consumer Electron,Vol.45, No.3, 1999, pp. 592-597.

[2] Y.-C. Peng, C.-K. Liang, H.-A. Chang, H. H. Chen, and C.-J. Kao, "Integration of image stabilizer with video codec for digital video cameras", in *Proc. Int. Symp. Circuits Syst.*, 2005, Vol.5, pp.4781–4784.

[3] M. Okade and P.K. Biswas, "Fast Video Stabilization in the Compressed Domain", *IEEE International Conference on Multimedia and Expo (ICME)*, 2012, pp. 1015-1020.

[4] H. H. Chen, C.-K. Liang, Y.-C. Peng and H.-A. Chang, "Integration of image stabilizer with video codec for digital video cameras", IEEE Trans. Video Technology, Vol. 17, No.7, 2007, pp. 801-813.

[5] L. Marcenaro, G. Vernazza and C. S. Regazzoni, "Image stabilization algorithms for video surveillance applications", *Proc. Int. Conf. on Image Processing.*, 2001, Vol.1, pp. 349-352.

[6] J. Zhou, H. He and D. Wan, "Video Stabilization and Completion Using Two Cameras", IEEE Transactions on circuit and systems for video technology, Vol. PP, No.99, 2011, pp. 1-1.

[7] M. Oshima, T. Hayashi, S. Fujioka, T. Inaji, H. Mitani, J. Kajino, K. Ikeda and K. Komoda, "VHS camcorder with electronic image stabilizer", *IEEE Trans. Consum. Electron*, Vol.35, No.4, 1989, pp. 749–758.

[8] K. Sato, S. Ishizuka, A. Nikami and M. Sato, "Control techniques for optical image stabilizing system", *IEEE Trans. Consum. Electron*,Vol. 39, No.3, 1993, pp. 461–466.

[9] S.-J. Ko, S.-H. Lee, S.-W. Jeon and E.-S. Kang, "Fast digital image stabilizer based on gray-coded bit-plane matching", *IEEE Transactions on Consumer Electronics*,Vol. 45, No.3, 1999, pp. 598– 603.

[10] A. Ertürk and S. Ertürk, "Two-Bit Transform for Binary Block Motion Estimation", IEEE Trans. Circuit Syst. Video Technol, Vol.15, No.7, 2005, pp. 938- 946.

[11] S. Ertürk, "Multiplication-free one-bit transform for low complexity block-based motion estimation", IEEE Sig. Proc. Letters, Vol.14, No.2, 2007, pp. 109-112.

[12] N.-J. Kim, H.-J. Lee and J.-B. Lee, "Probabilistic Global Motion Estimation Based on Laplacian Two-Bit Plane Matching for Fast Digital Image Stabilization", EURASIP Journal on Advances in Signal Processing, 2008.

[13] W.Nan, H. Xiaowei, W. Gang and Y. Zhonghu, "An approach of Electronic Stabilization Based on Binary image matching of Color Weight", 2nd International Asia Conference on Informatics in Control, Automation and robotics (CAR), 2010, Vol.2, pp. 155-158,.

[14] S. Battiato and G.Puglisi, "Fast block based local motion estimation for video stabilization", IEEE Computer Society Conference on Pattern Recognition Workshops (CVPRW), 2011, pp. 50-57.

[15] A.A. Amanatiadis, and I. Andreadis, "Digital Image Stabilization by Independent Component Analysis", IEEE. Instrumentation and Measurement, Vol.59, No.7, 2011, pp.1755-1763.

[16] A. A. Yeni and S. Ertürk, "Fast digital image stabilization using one bit transform based sub-image motion estimation", *IEEE Transactions on Consumer Electronics*, Vol.51, No.3, 2005, pp. 917– 921.

[17] S. Ertürk, "Digital Image Stabilization with Sub-Image Phase Correlation Based Global Motion Estimation", *IEEE Trans. Consumer Electron*, Vol.49, No.4, 2003, pp. 1320-1325.

[18] S. Ertürk, "Image sequence stabilization: Motion vector integration (MVI) versus frame position smoothing (FPS)", Proc. 2nd IEEE R8-EURASIP Symposium on Image and Signal Processing and Analysis, ISPA'01, 2011, pp. 266-271.

[19] T.-Han. Tsai, C.-Lun. Fang and H.-M. Chuang, "Design and Implementation of efficient Video Stabilization Engine Using Maximum a posteriori Estimation and Motion Energy Smoothing Approach", IEEE Transaction On Circuit and Systems For Video Technology, Vol.22, No.6, 2012, pp. 817-830.

[20] J. Zhu and B. Guo, "Fast Layered Bit-Plane Matching for Electronic Video Stabilization", International Conference on Multimedia and Signal Processing (CMSP), 2011, Vol.1, pp. 276-280.

[21] L. Dong-bok, C. Ick-hyun, S. Byung Cheol and L. Tae Hwan, "ROI-Based Video Stabilization Algorithm for Hand-Held Cameras", *IEEE International Conference on Multimedia and Expo Workshops(ICMEW)*, 2012, pp. 314-318.

[22] S. Erturk, "Translation, rotation and scale stabilization of image sequences", *IEE Electronics Letters*,Vol. 39, No.17, 2003, pp.1245-12462.

[23] N.A. Tsoligkas, S. Xalkiadis, X. Donglai and I. French, "A guide to digital image stabilization procedure- An overview", 18th International Conference on Systems signals and Image Processing (IWSSIP), 2011, pp. 1-4.

[24] J.M. Wang, H.P. Chou, S.W. Chen and C.S. Fuh, "Video stabilization for a hand-held camera based on 3D Motion model", 16th IEEE Internat*ional conference on Image Processing (ICIP), 2009*, pp. 3477-3480.
[25] O. Nestares, Y. Gat, H. Haussecker and I. Kozinsev, "Video stabilization to a global 3D frame of reference by fusing orientation sensor and image alignment data", 9th IEEE International symposium on Mixed and Augmented Reality (ISMAR), 2010, pp. 257-258.
[26] M. Mohammadi, M. Fathi and M. Soryani, "A new decoder side video stabilization using particle filter", 18th International Conference on Systems signals and Image Processing (IWSSIP), 2011, pp. 1-4.
[27] S.-H. Yang and F.-M. Jheng, "An adaptive image stabilization technique", In Proceedings of the *IEEE International Conference on Systems, Man, and Cybernetics* (SMC2006), 2006, Vol.3, pp. 1968-1973.
[28] F. Vella, A. Castorina, M. Mancuso and G. Messina, "Digital image stabilization by adaptive block motion vectors filtering", IEEE Trans. Consumer Electron, Vol.48, No.3, 2002, pp. 796-801.
[29] S. Battiato, G. Puglisi and A.R. Bruna, "A Robust Video Stabilization system by adaptive motion vectors filtering", IEEE International Conference on Multimedia and Expo, 2008, pp. 373-376.
[30] J.-P. Hsiao, C.-C. Hsu, T.-C. Shih, P.-L. Hsu, S.-S. Yeh and B.-C. Wang, "The Real-time Video Stabilization for the rescue Robot", ICROS-SICE International Joint conference, 2009, pp. 4364-4369.
[31] K. Uomori, A. Morimura and H. Ishii, "Electronic image stabilization system for *video* cameras and VCRs", *J. Soc. Motion Pict. Telev. Eng.*, Vol.101, 1992, pp. 66-75.
[32] S. Ertürk and T.J. Dennis, "Image sequence stabilization based on DFT filtering", *IEE Proc. on Image Vision and Signal Processing*, Vol.127, No.2, 2000, pp. 95-102.
[33] S. Ertürk, "Real-time digital image stabilization using Kalman filters", *Real-Time Imaging*, Vol.8, 2002, pp. 317-328.
[34] M.K. Güllü, E. Yaman and S. Ertürk, "Image sequence stabilization using fuzzy adaptive Kalman filtering", *Electronics Letters*, Vol.39, No.5, 2003, pp. 429-431.
[35] M.K. Güllü and S. Ertürk, "Fuzzy Image Sequence Stabilization", *Electronics Letters*, Vol. 39, No.16, 2003, pp. 1170-1172.
[36] M.K. Güllü and S. Ertürk, "Image sequence stabilization using membership selective fuzzy filtering", *Lecture Notes on Computer Science (LNCS)*, Vol.2869, 2003, pp. 497-504.
[37] M.K. Güllü and S. Ertürk, "Membership Function Adaptive Fuzzy Filter for Image Sequence Stabilization", *IEEE Trans. Consumer Electronics*, Vol.50, No.1, 2004, pp. 1-7.
[38] N. Kyriakoulis and A.Gasteratos, "A Recursive Fuzzy System for Efficient Digital Image Stabilization", *Hindawi Advances in Fuzzy Systems*, 2008.
[39] E. Yaman and S. Erturk, "Image Stabilization by Kalman Filtering Using a Constant Velocity Camera Model with Adaptive Process Noise", Proc. Of ELECO'2001, 2001, pp.152-155
[40] C. Wang, J-H. Kim, K-Y. Byun, J. Ni and S-J. Ko, "Robust Digital Image Stabilization Using the Kalman Filter", IEEE Transaction on Consumer Electronics, Vol.55, No.1, 2009, pp.6-14.
[41] B. Pinto and P.R. Anurenjan, "Video stabilization using speeded Up Robust features", International conference on communications and Signal processing (ICCSP), 2011, pp. 527-531.
[42] Road, http://www.jnack.com/adobe/photoshop/videostabilization/
[43] Shaky Car, matlabroot\toolbox\vipblks\vipdemos\shaky_car.avi

**Mohamad Javad Tanakian** was born in Zabol, Sistan and baluchestan, Iran in 1985. He received the B.S and also M.Sc degrees in Electrical engineering in 2007 and 2011 from university of Sistan and Baluchestan. His research interests are in the area of signal processing, Image and video processing and Fuzzy systems.

**Mehdi Rezaei** (IEEE M'04) received his B.Sc. degree in electronics engineering from Amir Kabir University of Technology (Polytechnic of Tehran) in 1992, his M.Sc. degree in electronics engineering from Tarbiat Modares University of Tehran in 1996, and his PhD degree in Signal Processing from Tampere University of Technology, Finland, in 2008. Dr. Rezaei had a closed cooperation with the Nokia Research Center as researcher and senior researcher from 2003 to 2009.
Dr. Rezaei is an Assistant Professor and also Deputy Dean for Education and Postgraduate Studies, in Faculty of Electrical and Computer Engineering, University of Sistan and Baluchestan, Iran. His research interests include multimedia signal processing and communications. He has published several journal and conference papers in these fields. He also holds several patents in related applications. Dr. Rezaei was the recipient of the Nokia Foundation Award in 2005 and also in 2006. He is a member of IEEE SP, COMSOC, and CAS Societies. Moreover, he is also a member of the Iranian computer vision and image processing society and also Iranian fuzzy systems society.

**Farahnaz Mohanna** received her B.Sc. degree in electronics engineering from University of Sistan and Baluchestan, Iran in 1987, her M.Sc. degree in electronics engineering from Tehran University in 1992, and her Ph.D. degree in Image Processing from Surrey University, Guilford, UK, in 2002. Dr. Mohanna has been working as a research fellow at the Centre for Vision, Speech and Signal Processing (CVSSP) at the Surrey University, UK in 2003. Dr. Mohanna is an Assistant Professor and Faculty Dean of Electrical and Computer Engineering, University of Sistan and Baluchestan, Iran. Her research interests include corner detection, active contours, object tracking, and video database retrieval and communications. She has published several Journal and conference papers in these fields. She is a member of the computer vision and image processing society.

# Camera Identification Algorithm Based on Sensor Pattern Noise Using Wavelet Transform, SVD / PCA and SVM Classifier

Kimia Bolouri*
Software Engineering, Payam Noor University, Tehran, Iran
bolouri_kimia@yahoo.com
Mehdi Javanmard
Electrical and Computer Engineering, Payam Noor University,Tehran,Iran
javanmard@pnu.ac.ir
Mohammad Firouzmand
Telecommunication Engineering, Iranian Research Organization for Science and Technology (IROST), Tehran, Iran
firouzmand@irost.org

Received: 02/Jun/2013 Accepted: 11/Dec/2013

**Abstract**
Identifying the source camera of an image is one of the most important issues of digital court and is useful in many applications, such as images that are presented in court as evidence. In many methods, the image noise characteristics, extraction of Sensor Pattern Noise and its correlation with non-uniformity of the light response (PNU) are used. In this paper we have presented a method based on photo response non uniformity (PRNU) that provides some features for classification by support vector machine (SVM). Because the noise model is affected by the complexity of the image, we used the wavelet transform to de-noise and reduce edge effects in PRNU noise pattern and also raise the detection accuracy. We also used the Precision processing theory to reduce the image size, then we simplified and summarized the data using the Single Value Decomposition (SVD) Or principal component analysis (PCA). The results show that using two-level wavelet transform and summarized data is more suitable using PCA.

**Keywords:** Sensor Pattern Noise, Camera Sensor Pattern Noise, Source Camera Identification, Photo Response Non-Uniformity, Wavelet Transform.

## 1. Introduction

Digital images can be taken by various types of digital cameras. In some applications, it is important to determine the source of a digital image. Since the digital images and videos can be easily fabricated, their contents are not reliable.

Thus identifying the source camera may be useful in judgments about images. Especially the electronic image detection techniques are important in the court. For example, identifying the source tool can determine the main images as evidence.

Identifying the source camera of an image is a complex issue that requires understanding the process of creating a graphic image of a real scene. In particular, it is necessary to know what is real and staged processes that affect the final digital data. In addition, it is necessary to consider the factors that may help to identify the camera. Many detection techniques have been proposed in the scientific literature. In general, most of these techniques work using the sensor noise (unwanted changes that occur in digital signal). This noise remains by digital sensor as a fingerprint in the image at the time of shooting.

In this paper, we first provide a summary of the performed activities, and then present our proposed algorithm. In the second section, we explain the processing stages in a digital camera that causes defects and noise in the image, especially the noise model. The proposed algorithm is described in the fourth section. Then, a detection technique is presented using support vector machine (SVM) that classifies the images into the corresponding camera. The efficiency of procedure is evaluated in the fifth section.

## 2. Related Works

Several methods have recently been proposed to detect camera. Reference [1] proposes an approach that detects the camera type using the noise correlation related to the Photo Non-uniformity (PNU). In this method, it is necessary to collect a large number of images related to a specific camera to take out an average of their residual noise, so that the camera Sensor Pattern Noise can be obtained. Because the image includes a wide range of noise and we have to nullify the effect of noises and just get the camera Sensor Pattern Noise, thus averaging enables us to achieve this goal. After that, a correlation threshold is used between the camera noise model and the image noise model to identify the source.

Reference [6] tests this method with changed images and the results show that in most cases this method is resistant against changes caused by the image processing functions. Reference [3,5] suggests a new idea for camera

* Corresponding Author

detection which is based on the collection of some image features. In this method, each image is presented as a vector of numerical features and the source camera can be detected by a classifier.

In some methods, it is necessary to consider some backgrounds (E.g. methods [9-12]), but the methods that are based on Sensor Pattern Noise do not require such assumptions.

The basic problem in the noise model extracted from image is that the extracted noise is not the real Sensor Pattern Noise, because the other random noises are affected by the image edges and complexity. Thus, reference [4] suggests some improvement for the Sensor Pattern Noise In this context, reference [7] finds some image retrieval methods however, this method is also correlation.

In this study, we use wavelet transform to reduce the complexity and the edges of the image. We also use SVD/PCA to reduce the data dimensions. Then, the obtained models are classified by SVM.

## 3. Digital Camera and Noise Types

In this section, we explain the processing stages that occur in a conventional digital camera and the noise types that enter the image at the time of shooting and are used as a fingerprint to identify the camera.

Light enters into a set of lenses and passes the counter -intuitive filters. Then it can be available to the color filter array (CFA) and considers one of the components of red (R), green (G) and blue (B) in each pixel because the sensor is only capable of leading one color at each pixel. Then an interpolation process is used to estimate the intensity of two other colors for each pixel using neighborhoods. A sequence of image processing operations such as color correction, white balance, gamma correction, image enhancement, and JPEG compression have been done, then the image is placed in the memory space. Some features of this process can be used to identify the type of camera as the sensor pattern noise (SPN), the response function of the camera, re-sampling, color filter array interpolation, JPEG compression and lens aberrations.

Some of these sources of noise are temporary and the others are based on distance and other types of noise are a combination of both.

Noise fluctuates over time indicate the "random" or "temporary" noise. There are three main types of temporary noises in optical and electronic systems: shot noise, thermal noise, and flicker noise. All of them can be observed in image sensors of CMOS and CCD. The noise which appears in the reconstructed image and is fixed in a certain place of the image implies a fixed pattern noise (FPN). Since it is fixed in terms of distance, then can be removed in the dark spots using the signal processing. Initial component (FPN) of in a CCD image sensor is the current non-uniformity of dark spots. This could be due to the time of shooting image or high temperatures. the main sources of FPN in CMOS image sensors are the current non-uniformity of dark spots And differences in the efficiency of an active transistor in a pixel. The important thing about FPN is that the spatial pattern of these changes remains constant. Since the FPN is added to all of the frames or images generated by a sensor and are independent of the light, it can easily be removed by subtracting a dark frame from the image.

A source which is somewhat similar to FPN in terms of properties is called the non-uniform response to light. One of its reasons is the non-uniform size of the active area through which light photons are absorbed. This effect is linear. PRNU is the result of Silicon material heterogeneity and effects that have been created throughout the manufacturing process.

In the next section we describe the proposed algorithm which is based on the use of PRNU noise.

## 4. Proposed Algorithm

Because there are different cameras with different features due to the use of sensors and other parameters, different models are derived from the images. Thus, this model could be used as a fingerprint of imaging devices to identify the source image. As mentioned before, our algorithm is based on PRNU pattern extraction and SVM classification. This algorithm has two parts: feature extraction and classification.

### 4.1 SVM Classification

Support Vector Machines (SVM) is a set of methods related to supervised learning that analyze the data and recognize patterns. Supervised learning is one of the machine learning methods to infer a function for supervised training data. To work with SVM, it is necessary to display each of the samples as a real numbers vector. Training data include a set of examples for training. We considered 30 images from each camera for training and tried to make various combinations of complexity or light in the image. In supervised learning, each sample has two components of inputs and outputs. Here the input component is a vector of qualities which is extracted from PRNU noise (which is different for each algorithm). Also we used PCA (principal component analysis) and SVD (single value decomposition) to reduce data size [11] and the most important data used in the models. Outputs are tags that define each of the four cameras. Supervised learning algorithm analyzes the training data and produces an inferred function that is called classifier. Inferred function forecasts the correct output value for any valid input component. It requires the Learning algorithm that is generated from the training data to cover the non-trained positions.

### 4.2 Features Extraction

The data used in this study consist of images of four cameras which are shown in Table 1 [14]. We considered 30 images for training and 100 images for testing per camera.

Table 1: Cameras that used for testing and training

| ID | Model | Sensor | Image Size |
|---|---|---|---|
| C1 | Canon EOs400D | CMOS | 3888×2592 |
| C2 | Kodak EasyShare CX7530 | CCD | 2560×1920 |
| C3 | HP PhotoSmart E327 | CCD | 2560×1920 |
| C4 | Panasonic DMC-FZ20 | CCD | 3648×2738 |

PRNU noise is extracted by de-noising filter that is based on wavelet transform. The overall goal of de-noising filter is described in [2]. This filter derives the Gaussian noise with a known variance (which enters the filter as an input), based on this assumption, in the wavelet domain, image and noise are an increasing combination from a non-stationary Gaussian noise that has a certain variance. The tests show that the value of 5 for the given variance has the best efficiency everywhere. We obtain the residual noise from the image by removing the de-noised image from the original image. (Equation 1)

$$\text{Residual Noise}=I-I_{denoised} \quad (1)$$

Here Idenoised is obtained with wavelet de-noise filter ([2]). Thus we obtain the high frequency components of the image that have been removed in the de-noised image as the image pattern noise. It is necessary to explain that the complexity and edges of the image influence the obtained pattern. You can see the result of implementing this method in PRNU noise extraction from a test image in Figure 1. In this method, PRNU is a matrix with image dimensions.

Fig 1: (left) original image, (right) PRNU noise pattern that includes high frequency components of the image.

In the most methods, the method of averaging the extracted noise pattern is used to obtain the camera noise model. This averaging is done for all images that were taken from a flat surface. Then, the camera type is detected using the correlation number of image noise model and camera noise model in cases such as [4,7] it has been tried to improve the extraction model.

The implementation results based on correlation can be observed in Table 2. In this method, vectors are the numbers related to the correlation rate of each image with the corresponding camera noise model in each color channel.

Table 2: Results of implementing algorithms based on averaging and correlation

| | Train Accuracy (30 images) | Test Accuracy (100 images) |
|---|---|---|
| C1 | 84.17% | 78.50% |
| C2 | 87.50% | 84.25% |
| C3 | 78.33% | 77.75% |
| C4 | 97.50% | 95.50% |
| Average | 86.88% | 84% |

Filtering process affects the pixel values around the edges of the image. So if the image has a complicated structure, its detection accuracy will be affected.

To solve this problem, we will propose an approach that is based on wavelet decomposition. Discrete Wavelet Transform (DWT) ([10]) divides the image into four sub-bands LL, HL, LH and HH (Figure 2).

We ignore the coefficients of HL, LH and HH that include vertical, horizontal and diagonal edges, then use The LL that represent the image approximate version where the edges have been trimmed. Furthermore, with this technique, we can reduce the size of image matrix and improve the efficiency PRNU noise extraction and vectors classification. Considering all the above issues, we use the LL part of wavelet transform and extract PRNU.

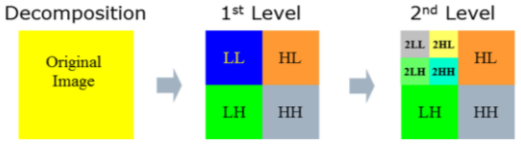



Figure 2: Bands of wavelet transform to the level 2

The purpose of using wavelet transform is to reduce the effects of edges. In Figure 3 the PRNU noise of original image and PRNU noise of LL part in wavelet transform in are compared. As you can see, the edges in the pattern obtained from the LL are trimmed, and the matrix size is reduced. However, this model can be used for SVM classification.

Steps that are used in the algorithm for each image to classify the input vectors are described below:

1. We calculate the wavelet transform of the image in the LL. (using Equation1)
2. We extract the Residual noise from the LL.
3. We use SVD or PCA to reduce the data size.
4. We consider a Vector obtained in the previous section as the input of SVM machines.

You can see the results for Level 1, 2 and 3 of wavelet transform in Table 3.

In this regard, dimensions of image become smaller 4l times in which 1 is equal to wavelet transform level. When we extract the residual noise from LL part of wavelet transform of the image in level 1, we have a matrix with quadrant of the image size, when we use SVD/PCA we reduce the dimension of this and has a vector (singular values in SVD and variances in PCA), Then we use 1000 features to classify. By a similar procedure for levels 2 and 3 of the wavelet transform, the number of the features is

500 and 200 to classify. So, we use 30 vectors that are extracted from 30 images for training SVM, then test the classifier with 100 vectors of test images.

By data reduction, we reduced arithmetic operation and increased detection rate significantly.

As you can see in the results, although the efficiency of this algorithm is good, the detection accuracy has also improved.

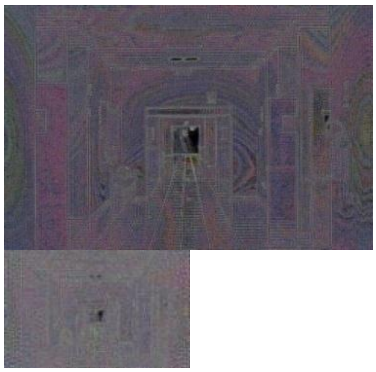


Figure 3: (Top) PRNU of original image, (bottom) PRNU of the LL in wavelet transform, edges are faded out and the image size is reduced.

## 5. Summary and Conclusions

In this paper, an algorithm is proposed for digital camera identification. The previous methods used the averaging of noise pattern to reduce the effects of image complexity on the noise model. This method is very time consuming and requires the purchase of a larger database containing images of a smooth surface for averaging. According to the results, the numbers obtained from correlation are not a suitable option for SVM classification. In cases such as [4], which some ameliorative are used for the noise model, the average accuracy of 80.8% has been achieved. In the proposed method, to reduce the complexity and edges of image and improve the noise model, the LL part of wavelet transform is used. It is clear that if the level is higher, the speed will increase due to the image size reduction but the algorithm accuracy will decrease. The results indicate that the PRNU extraction from the LL part of wavelet transform will give the best results in level 1.

Table 3: Results of using wavelet transform to fade out the edges of the image noise pattern

| | Train Accuracy(30 images) | | | | | | Test Accuracy(100 images) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DWT Level | 1 | | 2 | | 3 | | 1 | | 2 | | 3 | |
| Data Reduction | SVD | PCA | SVD | PCA | SVD | PCA | SVD | PCA | SVD | PCA | SVD | PCA |
| C1 | 99.17% | 98.33 | 100% | 99.2% | 100% | 98.3% | 96% | 91.75 | 97.5% | 96.8% | 96% | 92% |
| C2 | 91.67% | 92.5% | 88.3% | 92.5% | 87.5% | 90% | 85.75% | 81.25 | 83.8% | 88.3% | 79% | 85.5% |
| C3 | 95.83% | 96.67% | 95.8% | 95% | 93.3% | 93.3% | 92.25% | 92.75% | 81.3% | 80.3% | 83% | 86% |
| C4 | 98.33% | 99.17% | 100% | 100% | 100% | 100% | 92.5% | 94% | 97.3% | 94% | 97.7% | 95% |
| Average | 96.25% | 96.67% | 96.03% | 96.68% | 95.20% | 95.40% | 91.63% | 89.94% | 89.94% | 89.85% | 88.93% | 89.63% |

## References


[1] Lukáš Jan, Fridrich Jessica, and Goljan Miroslav, "Digital Camera Identification from Sensor Pattern Noise ", IEEE Signal Processing Society, Volum: 1, Issue: 2, June 2006, p.205-214 .

[2] M., Igor Kozintsev, and Kannan Ramchandran. "Spatially Adaptive Statistical Modeling Of Wavelet Image Coeffcients And Its Application To Denoising", Acoustics, speech, and Signal Processing, Volume: 6, 15-19 Mar 1999, p:3253-3256.

[3] Kharrazi Mehdi, Husrev T, Sencar b, Nasir Memon, "Blind Source Camera Identification", Image Processing, ICIP, Volume: 3, Sept 2005, p.69-72.

[4] Chang-Tsun Li, "Source Camera Linking Using Enhanced Sensor Pattern Noise Extracted from Images", Information Forensics and security, Volum: 5, Issue: 2, June 2010, p.280-287.

[5] Nitin Khannaa, Aravind K. Mikkilinenib,George T. C. Chiub, Jan P. Allebacha, Edward J. Delpa, "Scanner Identification Using Feature- Based Processing and Analysis", Information Forensics and Security, Volume: 4, Issue: 1, March 2009, p:123-139.

[6] Castiglione Aniello, Cattaneo Giuseppe, Cembalo Maurizio, "Source Camera Identification in Real Practice: a Preliminary Experimentation", Broadband, Wireless Computing, Commiunication and Application, IEEE Computer Society, 2010, p:417-422.

[7] Kazuya Matsushita and 2Hitoshi Kitazawa," An Improved Camera Identification Method based on the Texture Complexity and the Image Restoration", International Conference on Hybrid Information Technology, ICHIT, ACM, 2010, p.171-175.

[8] Sevinc Bayram a, Husrev T. Sencar b, Nasir Memon b, Ismail Avcibas, "Source camera identification based on CFA interpolation", IEEE Transaction on Information Forensics and Security, Volume: 4, Issue: 4, December 2009.

[9] Chang- Hee- Choi, Jung- Ho choi, Heung- Kyu Lee, "CFA pattern identification of digital cameras using intermediate value counting", MM&Sec'11, Processing of thirteenth ACM multimedia workshop on multimedia and security, 2011, p.21-26.

[10] Gonzales R and Woods E, "Digital Image Processing", wavelet and multiresolution processing (c-7), 3nd edition, Prentice Hall, 2008, pp.449-511.
[11] "A Tutorial on Principal Component Analysis", version 3.01, Systems Neurobiology Laboratory, Salk Insitute for Biological Studies, Shlen, New york, 2009, April.
[12] Markowetz Florian, "Classification by Support Vector Machines", Florian Markowetz Max-Planck- Institute for Molecular Genetics, Computational Molecular Biology, Berlin ,2003
[13] Burges Chistopher J.C. "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery 2, 1998, pp.121-167.
[14] Data set of images to enrollment, train and test for camera identification http://acqua.dia.unisa.it/DID, March 2012.

**Kimia Bolouri** received her B.S degree in software engineering from Shiraz Azad University, Shiraz, Iran, in 2003 and received M.S degree in Tehran Payame Noor University, Iran in 2013. Her research interests include image processing, camera identification algorithms, She works at IT organization of Shiraz municipality and analyzes, designs and develops several projects and applications.

**Mahdi Javanmard** received his M.Sc. degree in Electrical Engineering from the University of New Brunswick, Canada and Ph.D. degree in Electrical and Computer Engineering from Queen's University at Kingston, Canada, in 1989 and 1996 respectively. He is a faculty member of Payam Noor University (PNU) and currently Head of COMSTECH Inter Islamic Network on Virtual Universities (CINVU). He has been teaching for many years at different universities where he has been involved in their course development for the Computer Science Department. Additionally, he works as a System Development Consultant for various companies. Dr. Javanmard's research interests are in the areas of Information & Communication Security, Speech Recognition, Signal Processing, Urban Management & ICT, and Ultrasound Medical Imaging. For more information, please see javanmard.com.

**Dr. Mohammad Firouzmand** received his M.S. degree in Bio_Medical Engineering in Sharif Univ. in 1992 and his Ph.D. degree in signal processing in Institute National Polytechnique de Grenoble (INPG), France, in 2007. He is currently an Associate Professor in Iranian Research Organization for Science and technology (IROST). His research interests include Bio-instruments, signal processing, image and audio signal processing, and he has published several research papers in these areas.

# Video Transmission Using New Adaptive Modulation and Coding Scheme in OFDM based Cognitive Radio

Hasan Farsi*
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.
hfarsi@birjand.ac.ir
Farid Jafarian
Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.
farid.jafarian@birjand.ac.ir

Received: 01/Jul/2013 Accepted: 11/Dec/2013

**Abstract**
As Cognitive Radio (CR) used in video applications, user-comprehended video quality practiced by secondary users is an important metric to judge effectiveness of CR technologies. We propose a new adaptive modulation and coding (AMC) scheme for CR, which is OFDM based system that is compliant with the IEEE.802.16. The proposed CR alters its modulation and coding rate to provide high quality system. In this scheme, CR using its ability to consciousness of various parameters including knowledge of the white holes in the channel spectrum via channel sensing, SNR, carrier to interference and noise ratio (CINR), and Modulation order Product code Rate (MPR) selects an optimum modulation and coding rate. In this scheme, we model the AMC function using Artificial Neural Network (ANN). Since AMC is naturally a non-liner function, ANN is selected to model this function. In order to achieve more accurate model, Genetic algorithm (GA) and Particle Swarm Optimization (PSO) are selected to optimize the function representing relationship between inputs and outputs of ANN, i.e., AMC model. Inputs of ANN are CR knowledge parameters, and the outputs are modulation type and coding rate. Presenting a perfect AMC model is advantage of this scheme because of considering all impressive parameters including CINR, available bandwidth, SNR and MPR to select optimum modulation and coding rate. Also, we show that in this application, GA rather than PSO is better choice for optimization algorithm.

**Keywords:** Adaptive Modulation And Coding; Cognitive Radio; IEEE 802.16 Standard; Video Transmission; Wireless Channel.

## 1. Introduction

Today's with the growth of wireless systems, there is an expanding demand on real-time wireless video communications. However, transmission over Wireless channels leads to a complicated problem due to the multipath fading behaviors of the channel [1], [2] and [3]. Cognitive radio (CR) is schemed at better resource management and ameliorated data transmission technologies. Through these purposes, it is designed to amalgamate with both the introduction of Software Defined Radios (SDR) and the consciousness that machine learning. CR can be defined as a system consisting of flexible wireless system design, measurements and awareness of various parameters, including interference temperature and geo-location information [4], [5].

An Adaptive modulation and coding (AMC) can be proposed as an effective part to the CR system. AMC is a technique that selects an appropriate pair of a modulation and a channel coding rate. In [6] a Fuzzy logic based CR is proposed. In this study, using Fuzzy logic analysis, the cognitive engine controls the form of modulation and code rate in order to take care the system throughput. In [7], an AMC method for selecting the appropriate Modulation and Coding Scheme (MCS) according to the estimated channel condition for 3G wireless system is proposed. There are several schemes for AMC, which can be found in [8-14]. In [15], an AMC scheme is proposed according to the variable power and variable-rate technique from [16] and [17]. This technique lays over a trellis code on top of the un-coded modulation.

In [18] an AMC technique is proposed using M-array phase shift keying (MPSK) modulation. Recently a resource allocation scheme is proposed for OFDMA based cognitive radio for video transmission [19], this paper performs subcarrier, bit, and power allocation for diferent cognitive users such that the sum rate of the cognitive users is increased, towards this aim, fine grain scalable (FGS) video is employed. Although this paper is near-optimal allocation scheme, it is not concern with channel coding which is inseparable part of communications systems. In the other words, channel code rate and its related ability to correct the channel errors, and also its impact on the bit-rate is not considered in optimization problem and simulation procedure. Although all of introduced AMC methods and resource allocation schemes provide acceptable performance, they are not perfect schemes for efficient CR because these schemes have not considered all involving parameters such as available bandwidth, SNR, carrier to interference and noise ratio (CINR), and Modulation order product

* Corresponding Author

code Rate (MPR), jointly, to select modulation and coding rate. Since CR, system requires a perfect scheme for AMC, in this paper, we present a perfect AMC scheme for OFDM based CR according to IEEE 802.16 [20], [21], that consider all the involving parameters. As mentioned, because of involving several parameters to determine the perfect AMC, behavior of such a function is completely nonlinear. Therefore to describe this function we propose employing the powerful Artificial Neural Network (ANN). Advantage of this scheme is considering parameters including CINR, available bandwidth, SNR, and MPR to select optimum modulation and coding rate. Consequently, the system presents a perfect and powerful decision to select optimum modulation and coding rate. This paper is organized as follows: in section 2 and 3, a briefly of OFDM based CR and H.264 video coding are presented respectively. In section 4 two recent AMC model are summarized. In section 5, our proposed scheme is detailed. Finally in section 6, our simulation setting is introduced and in section 7 simulation results are presented.

## 2. OFDM Based Cognitive Radio

In the last years, researchers have established an efficient inter-relation between software and radio systems. This has permitted for faster advancements and has provided wireless communication instruments more flexibility and the capacity to transmit and receive using a multiplicity of protocols. Mitola in [4] captured the definition of an SDR one step further and pictured a radio which could take decisions to the network, modulation and coding parameters based on its surroundings, and called as "smart" radio the CR. Implementation of OFDM into CR comes up as a new aspect and challenges to system design.

The cognitive engine is responsible for making the intelligent decisions and configuring the radio. The spectral opportunities are recognized by the decision unit based on the information from policy engine as well as local and network spectrum sensing data. The local spectrum sensing unit processes spectrum knowledge and then recognizes certified users accessing the spectrum and their signal setting such as bandwidth and power level, and disclose spectrum opportunities that can be utilized by CR. CR decisions contain selecting the suitable channel coding, modulation, operation frequencies, and bandwidth. At this stage, OFDM technology gets the best performance over other similar transmission technologies with its flexible features. By only changing the configuration parameters of OFDM, it can optimize the transmission relying on the environmental characteristics.

## 3. H.264 Video Coding Paradigm

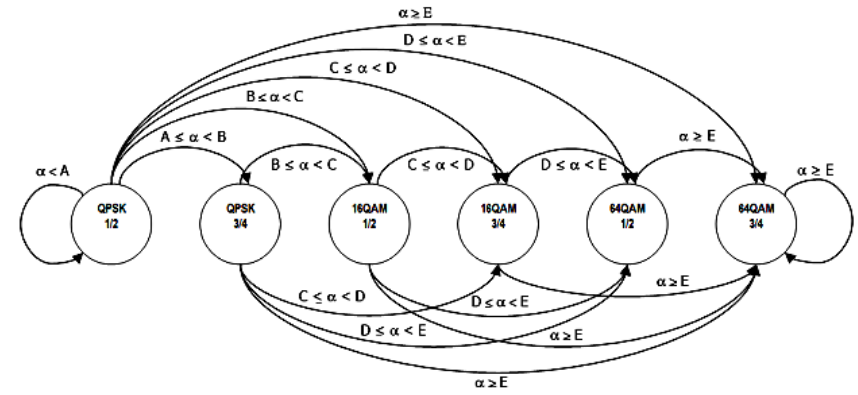
The joint project of ITU-T and ISO/IEC leads to developing the video coding design with the high coding efficiency which is called H.264/ advanced video coding (AVC) [22]. The primary goal of designing the H.264/AVC is to develop a simple and genuine video coding structure with improved compression performance. Currently, it becomes the standard of video coding because of the best performance in terms of its rate-distortion compared with the previous standards in the wireless networks.

In H.264, a set of video frames are transmitted in which each frame is partitioned to several macro-blocks. Each macro block (MB) comprises a motion vector (MV). The coding process of the first frame (I frame) of a video sequence is called "intra" coded in which 2D-DCT is applied on the MBs. Then the resulting coefficients are quantized and transmitted. The coding process of other remaining frames (P frames) of a video sequence, is called "inter" coded. Inter coding uses motion compensation which is the prediction of the current frames based on previous frame by employing motion estimation (ME) part. The Inter prediction encoding process consists of ME and providing the motion data transmitted as side information. Next, 2D-DCT is applied on the residual of the prediction (error of prediction) which is the difference between the original and the predicted block. Then resulting coefficients are quantized and transmitted. Finally, the quantized transform coefficients are entropy coded and transmitted together with the side information. The high coding efficiency of H.264 can be achieved by advanced coding tools such as multiple reference frames, variable block size and quarter-pixel accuracy ME [23].

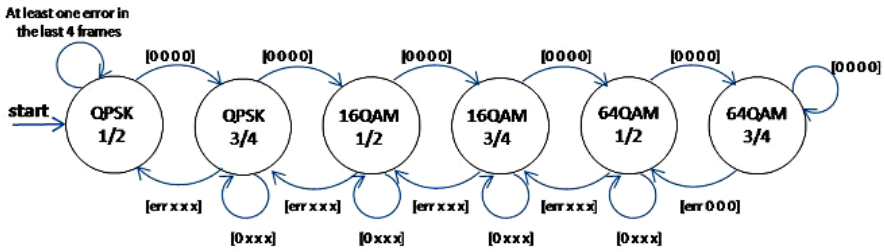
## 4. Recent AMC Schemes

AMC schemes are widely considered and investigated in wireless communication systems in the recent literatures [6-19]. In this paper, we concentrate on two efficient AMC schemes presented in [30]. These AMC schemes are called Channel State scheme and Error state scheme. These schemes are the only recent benchmarks possible which are based on a slot allocation.

Channel State scheme is designed to adapt the modulation and coding scheme (MCS) after deriving an estimation of the channel behavior in terms of attenuation coefficient, $\alpha$.




(a) Moore's state machine for channel state techniques.




(b) Moore's state machine for error techniques.

Fig.1. Moore's state machines: 0=correct frame, err=error frame and x=any frame [30].

Over a slot, attenuation coefficient is compared with suitable thresholds, due to determination of the MCS to be used in the next frame. In Fig. 1a, Moore's state machine of this scheme is shown for the case of five thresholds A, B, C, D and E where A is the lowest threshold value [30]. In this scheme, two algorithms named Maximum Throughput (MT) and Target BLER have been proposed. The MT algorithm is designed to maximize the overall link throughput by selecting the proper MCS for each SNR value. As a result, this algorithm is not suitable for such services that are sensitive to the error probability [30], e.g., our application, real-time video communication. The other algorithm is TBLER designed based on keeping the error rate below a target constraint. This algorithm is more suitable for such services that are sensitive to the error probability. More details can be found in [30].

In the Error state scheme, the MCS changes in function of the number of detected frames with errors. In Fig. 1b, Moore's state machine of this scheme is shown, under the assumption of an error depth equal to four frames [30]. In this algorithm, transitions are permitted only for neighbor states (e.g., from QPSK 3/4 to 16QAM 1/2 and backwards but not from QPSK 3/4 to 16QAM 3/4). The transitions observe the following rules [30]:

1) A transition to a more efficient state happens only if all the last four frames are without errors (when we are the most efficient state (i.e., 64QAM 3/4) we still remain in that state).
2) A transition to less efficient state happens only if the last frame has at least one error (when we are in the less efficient state (i.e., 4QAM 1/2) we still remain in that state).
3) The state remains the same if the last frame is without errors independently from the previous three.

## 5. The Proposed Scheme

Since our proposed AMC scheme is based on population algorithms and Artificial Neural Network (ANN), we present a brief description of them and their application in our proposed scheme.

### 5.1 Population Based Algorithm

#### 5.1.1 Genetic Algorithm

GA as a stochastic optimization technique generates solutions to optimization problems using methods such as mutation, crossover, selection, and inheritance inspired by natural evolution. In this kind of techniques; information of each generation is passed on to next generation by chromosome. Each chromosome includes gens; also, each gen demonstrates a particular characteristic or manners [24]. In GA's method, the primary populations are firstly produced based on necessities of the problem, following that, objective function is estimated and in order to achieve best solution, in regeneration step, parents create children. In this step, some activities are happening such as crossover and mutation. Therefore, the best solution is obtained during determination and essential iterations.

#### 5.1.2 Particle Swarm Optimization (PSO)

PSO simulates the behaviors of bird gathering. In the PSO algorithm, each single solution is considered as a "particle" in the search space. Fitness value computes for all the p particles. The p particles are "flown" by way of the problem space by following the present optimum p particles. PSO is initialized by a group of arbitrary p particles (as we know, particles are solutions). Then, it investigates for optimal solution by renewing generations. In each iteration, each p particle is renewed by following two "best" values. The first one is the best solution which it has been obtained so far. This value is well known as "pbest". Another "best" value that is tracked by the p particle swarm optimizer is the best value obtained so far by any particle in the population. This best value is a global best called "gbest" [26-28].

After discovering the two best values, the particle renews its velocity and positions based on the following equations.

$$V_{new} = X \times (V_{old} + \phi_1 \times (P_{Pb} - X_{cs}) + \phi_2 \times (P_{gb} - X_{cs})) \quad (1)$$

$$X_{new} = X_{old} + V_{new} \quad (2)$$

$$X = \frac{2k}{\varphi - 2 + \sqrt{\varphi^2 - 4\varphi}} \quad for \ \varphi > 4, \quad k \in (0,1) \quad (3)$$

Where, $\phi_1$ and $\phi_2$ are random numbers uniformly distributed in the range (0, $\phi/2$) and $X$ is the constriction coefficient.

### 5.2 Artificial Neural Network (ANN)

ANNs are one of the most influential branches of artificial intelligence. Generally, it is used for description of non-linear functions, like the function of this paper. ANNs are categorized into two major groups based on learning algorithm, supervised and unsupervised learning. In supervised learning, input and desired output are presented as learning data and ANN is trained using them. In unsupervised learning, target outputs are not present and ANN can only cluster input data and when new data input, it can assign it into the corresponded cluster [25].

#### 5.2.1 Neural Network Structure

The neural network used in this paper is a multilayer perceptron neural network. The weights and biases are the variables of the optimization algorithm. Firstly, matrix of weights and biases indicating initial position of particles for PSO and chromosome for GA are prepared as follow:

$$Weight = [w_1, w_2, ..., w_i], \ Biase = [b_1, b_2, ..., b_i] \quad (4)$$

Where, w denotes weights and b denotes biases. Also for structure of ANN with 2 hidden layers we have:

$$\underline{f_1} = \underline{W_1} \times X + b_1 \quad (5)$$

$$\underline{\sigma_1} = \log sig(\underline{f_1}) \quad (6)$$

$$\underline{f_2} = \underline{W_{L2}} \times \underline{\sigma_1} + b_2 \quad (7)$$

$$\underline{\sigma_2} = \log sig(\underline{f_2}) \quad (8)$$

$$\underline{f_3} = \underline{W_{L3}} \times \underline{\sigma_2} + b_3 \quad (9)$$

$$\underline{\sigma_3} = \log sig(\underline{f_3}) \quad (10)$$

Where, $\underline{X}$ is vector of input data $\underline{W_{L1}}$, $\underline{W_{L2}}$ and $\underline{W_{L3}}$ are weights between input and first layer, between first layer and second one and between second layer and third one, respectively. b1, b2 and b3 are biases related to first layer, second layer and third layer, respectively. Finally, $\underline{\sigma_3}$ is actual output of network and MSE is value of subtraction of actual output and target output. This value is fitness value of GA and PSO, therefore the proposed algorithm will change weights until MSE becomes minimum in next iterations. MSE function is given by Eq. (11)

$$MSE = \frac{1}{L}\sum_{i=0}^{L-1}\left(Y_{output\_ANN} - Y_{target\_output}\right)^2 \quad (11)$$

Where, $Y_{output\_ANN}$ is the out-put of model and $Y_{target\_output}$ is expected out-put of the model. There is different methods to train (or in other words, to update weights and biases) which most of them employ analytical and mathematical based methods, such as back propagation and gradient descent. In this paper, to optimize the function representing relationship between inputs and outputs of neural network GA and PSO are used to optimize the function.

### 5.3 The Proposed AMC Model

In this scheme for cognitive Radio, at first, cognitive engine senses the spectrum using spectrum sensing unit [5]. Then it finds the white hole of spectrum. Next, because of its ability to awareness of changing in transmission environment, CR has knowledge of SNR, carrier to noise ratio (CINR), carrier to interference ratio (CIR), and Modulation order Product code Rate (MPR). Based on these parameters, it selects an optimum modulation and coding rate. We model the AMC function using ANN. Since AMC is naturally a non-liner function, ANN is selected to model this function. In order to achieve more accurate model, GA and PSO are selected to optimize the function representing relationship between inputs and outputs of ANN, i.e., AMC model. Inputs of ANN are CR knowledge parameters and the outputs are modulation type and coding rate. These parameters are calculated in following subsections.

#### 5.3.1 Modulation Order Product Code Rate (MPR)

For DL/UL, the transmission style is precondition by the NEP (encoding packet size) and the NSCH (number of allotted slots). NEP per an encoding packet is {144, 192, 288, 384, 480, 960, 1920, 2880, 3840, and 4800}. The NSCH per an encoding Packet is {1, 480}. In Table 507, 509 [29], the numbers in the first row are NEPs, and the numbers in the remaining rows are NSCHs and associated parameters. In Table 507,509 [29] the modulation order is desirable by MOD, and it has the values of 2 for QPSK, 4 for 16-QAM, and 6 for 64-QAM. SCH points to the number of allocated slots. The bearable modulation schemes are QPSK, 16-QAM, and 64-QAM.

When the NEP and the NSCH are given, the modulation order is decided by the value of MPR (Modulation order Product code Rate).

The MPR means the effective number of the information bit transmitted bits per a subcarrier and is defined by (12).

$$MPR = \frac{N_{EP}}{48 \times N_{SCH}} \tag{12}$$

### 5.3.2 Carrier to Interference and Noise Ratio (CINR)

One imaginable technique to estimate the CINR of a particular message is to normalize the mean-squared residual error of disclosed data symbols (and/or pilot symbols) by the average signal power [29] using:

$$CINR[K] = \frac{\sum_{n=0}^{N-1} |s[k,n]|^2}{\sum_{n=0}^{N-1} |r[k,n] - s[k,n]|^2} \tag{13}$$

CINR[k] is the (linear) CINR for message k, s[k, n] is the corresponding detected more pilot symbol corresponding to received symbol n, and r[k, n] is the received symbol n within message k. As mentioned in previous sections, ANN can be used as a suitable model for complex processes like AMC that depends on several parameters, such as CINR, SNR and bandwidth of white hole.

## 5.4 Optimization Problem: Optimization of Weights and Biases by GA and PSO

As mentioned before, GA has some parameters, which should be set in initialization step, such as Population Size, Elite Count, Migration Fraction, Migration Interval, crossover, and mutation. In this method the ANN is used as an objective function in optimization. The optimization algorithm is summarized in Algorithm 1.

**Algorithm 1:** Optimization of AMC model

Input data: weights and biases of ANN

1) Determining parameter of GA and PSO.
2) Determining initial position and velocity randomly.
3) Evaluating fitness of chromosomes /particles.
4) Generating new chromosomes/Updating particles velocity and position.
5) Checking the stop criterion:
   If the stop criterion is met the optimum AMC model is obtained, if not, go to step 2.

Output: Optimum AMC model.

Each chromosomes of the GA represents a set of weights and biases of one ANN which may separately be an answer to the optimization problem. Depending on the adjustments, these chromosomes are transformed into new chromosomes evaluated based on the objective function. In PSO, the procedure is similar to GA optimization and the only difference is that the population is particle. For a multilayer perceptron ANN with one hidden layer that consists of n hidden unit and the network has m inputs then the output of the network can be formalized as a function of the inputs. Let's define Y as the output of the ANN. The mathematical model of such a network defined as follow:

$$Y_{output-ANN}(x_1, x_2, ..., x_m) = \sum_{j=1}^{n} w_j \, f\left(\sum_{i=1}^{m} a_{ij} \, x_i + b_j\right) \tag{14}$$

Where $W_j$ is the weight of the synapse that goes to the output neuron form the j-th neuron and $a_{ij}$ is the weight of the synapse that goes form the input $x_i$ to the j-th hidden neuron, $b_j$ is the bias of the j-th hidden neuron, $f$ is the activation function. Optimization problem is defined as follow:

$$\min\left\{\frac{1}{L}\sum_{i=0}^{L-1}\left(Y_{output\_ANN} - Y_{target}\right)^2\right\} \tag{15}$$

To stop the optimization process, the amount of error in training the ANN is considered to be less than 0.3%. The optimization process continues until the above condition is satisfied.

## 5.5 The Structure of Proposed AMC Scheme and Optimization Results

In this paper, we use Multilayer perceptron ANN for finding the function of AMC. Based on the five input parameters (UL/DL, SNR, CINR, available Bit-Rate, MPR) and two outputs (type of modulation and channel code rate), we utilize an ANN which has five inputs, two outputs, and 2 hidden layers. First layer has 10 neurons and second layer has 8 neurons. The modulation order (1 for BPSK, 2 for QPSK, 4 for 16-QAM, and 6 for 64-QAM) will be set for all the allowed transmission formats as shown in Table 507 of [29] (Note in this paper we consider BPSK modulation for low SNR transmission environments). For training the system, we use Table 507-Transmission format and modulation level for DL, [29], Table 509-Transmission format and modulation level for UL, [29], Table B.4 OFDM 256-FFT raw bit-rates (Mb/s), [29] CNR is the normalized Carrier to Noise Ratio (per sub-carrier) for the given modulation, Table-308 of [29]. We combine these tables into two tables. The resulting tables are given in Tables 2 and 3. The Table 2, is for uplink transmission and the Table 3, is for downlink transmission. As mentioned in previous sections, ANN can be used as a suitable model for complex processes. Since AMC involves nonlinear and complicated features (the function representing AMC has multi-input variables and multi-output variables), accurate prediction of the behavior of this process is great significance. In this paper, the proposed method has three major steps. In first step, the function which represents the model of AMC is presented, in the second step, trained ANN with GA has been used to presents an efficient and optimum model for AMC. In third step, trained ANN with PSO has been used to present an efficient and optimum model for AMC. We use Table 2 and Table 3 to train and test the ANN, ANN-GA and ANN-PSO.

80 percent of this data table is used for training and 20 percent is used for test. In this paper we employ multilayer perceptron ANN which has five inputs, two outputs, and 2 hidden layers. First layer has 10 neurons

and second has 8 neurons. It should be mentioned that in our scheme the input parameters of ANN have the same importance. GA parameters set as; maximum number of iterations, Max-It=100; population size, N-pop=50; crossover percentage, P-crossover = 0.7; number of parents is calculated by following formula:

N-crossover = round (P-crossover * N-pop / 2) * 2, mutation percentage, P-mutation = 0.2; number of mutants, N-mutation = round (P-mutation*N-pop). PSO parameters set as; maximum number of iterations Max-It=100; swarm population size, N-pop=50; $\phi_1 = 2.05$; $\phi_2 = 2.05$. Based on the presented settings the training procedure of ANN, ANN-GA and ANN-PSO is performed. The simulation results are shown in Fig.2, Fig.3 and Fig.4 respectively. Fig.2 shows training procedure of ANN. It achieves the best validation in 0.0053 in epoch 8, in this epoch MSE train is 0.0031. Next, we use GA and PSO to optimize the ANN. As shown in Fig.3, after 100 generation, GA optimizes the ANN and, as a result, it achieves the best in 0.001 (MSE=0.001). Also as shown in Fig.4, PSO achieves the best in 0.0022 (MSE=0.0022). Finally, resulting mean errors have been reported in Table 1. It is observed that ANN-GA is the optimum algorithm for representation of function of AMC because of its lower minimum mean error

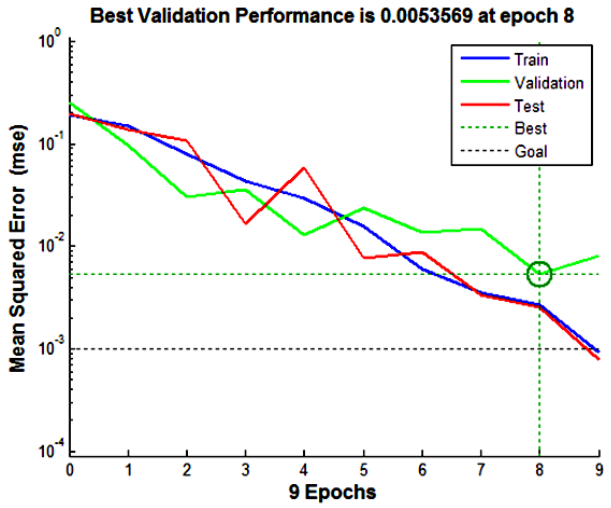



Fig.2. ANN simulation for AMC

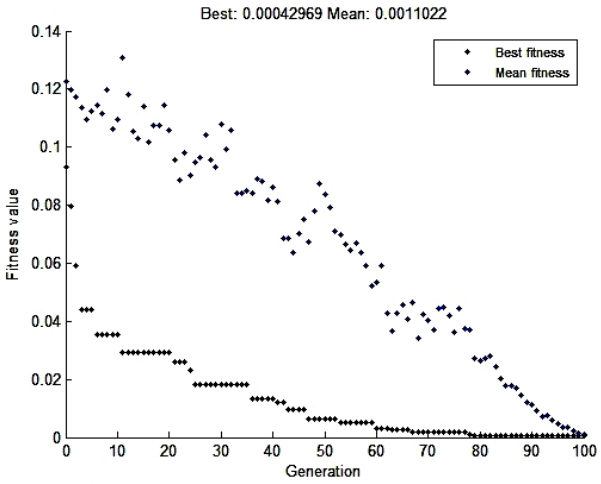



Fig.3. Optimized ANN simulation for AMC using GA

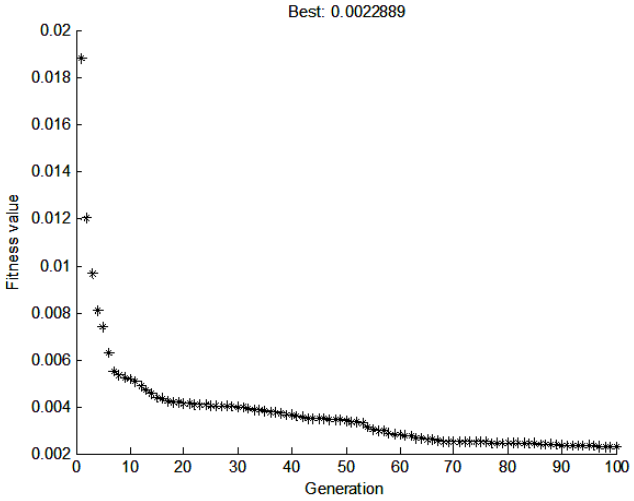



Fig.4. Optimized ANN simulation for AMC using PSO

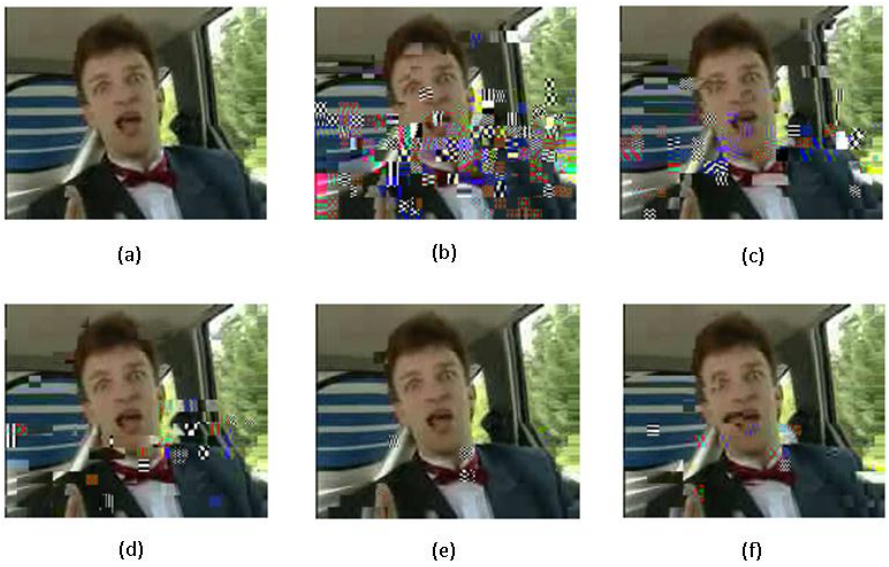



Fig. 5: Frame of Car-phone video in the receiver after passing Rayleigh fading channel in Uplink, SNR=10.5, MPR=1.5, CINR=23.7, (a) is the coded frame 285, (b) using QAM64 and Code rate = 0.25, (c) using QAM16 and Code rate = 0.5, (d) using QAM and Code rate = 0.67, (e) QAM16 and Code rate= 0.38 and (f) using QAM and Code rate = 0.63.

### 5.6 About the Complexity of Proposed Scheme

There may be several numbers of algorithms to solve the same function. In terms of complexity theory, one strategy to evaluate them is to compare their performance in terms of how quickly they solve the same function. In this paper, we emphasize that the function having been considered for AMC in other literatures is imperfect because this function does not consider important parameter such as CINR, MPR and available bandwidth (i.e. they consider a simple function for AMC and design algorithms to solve it). In this paper, we present a non-linear function for AMC including all impressive parameters; also, we design a scheme according to this complex function. Since we change the AMC function, clearly, we do not deal with the same function to compare the complexity of our scheme with the other schemes presented in other literatures.

Furthermore, the proposed AMC model and its optimization algorithm (Algorithm 1) are the pre-computing procedures which are performed offline, i.e., these procedures are applied to determine the function representing AMC model. The resulting AMC model can be applied in each cognitive radio system. What the system needs to compute online is $Y_{AMC}$, i.e., the system just enters the inputs, $x_i$, in equation (16) which its parameters $w_k$, $w_j$, $a_{ij}$ and $b_j$ are offline prepared already. $Y_{AMC}$ is our proposed AMC model (our model has 5 inputs, 2 outputs and 2 hidden layers):

$$Y_{AMC}(x_1, x_2, ..., x_m) = \sum_{k=1}^{h} w_k \; f_2\left(\sum_{j=1}^{n} w_j \; f_1\left(\sum_{i=1}^{m} a_{ij}\, x_i + b_j\right)\right) \quad (16)$$

As mentioned before, $Y_{AMC}$ is the AMC model. Therefore, the complexity for our proposed AMC scheme is just related to calculating the $Y_{AMC}$, i.e., $O(Y_{AMC})$.

Maybe this question arises why we apply ANN in order to model the AMC function. The reason is that we require a powerful tool that can be able to model a complex and nonlinear function which has several inputs and several outputs. ANN and nonlinear regression are known as powerful tools for modeling the complex and nonlinear functions. Although nonlinear regression is a powerful tool, it is proved that ANN is more accurate than it [32]; furthermore, it is a traditional tool respect to ANN.

## 6. Simulation Settings

Conventional OFDM system with assuming 256 subcarriers is used in our simulation. In wireless transmission, motion causes Doppler shift in the received signal components, the Doppler frequency $f_D$ equals to $f_c v/c$ in which v is the mobile speed, $f_c$ is the carrier frequency and c is the light speed and equals $3 \times 10^8$. In deriving our simulations results we have focused on a wireless environment with Radio frequency carrier $f_c = 3.5$GHz Maximum available bandwidth of 10 MHz, maximum Doppler frequency $f_D = 291$ Hz (mobility terminals up to 90 km/ h). Coherence time $T_c$ 1.03 ms. maximum delay spread equal to 20μs ITU-R vehicular channel model A, with 6 paths [33].

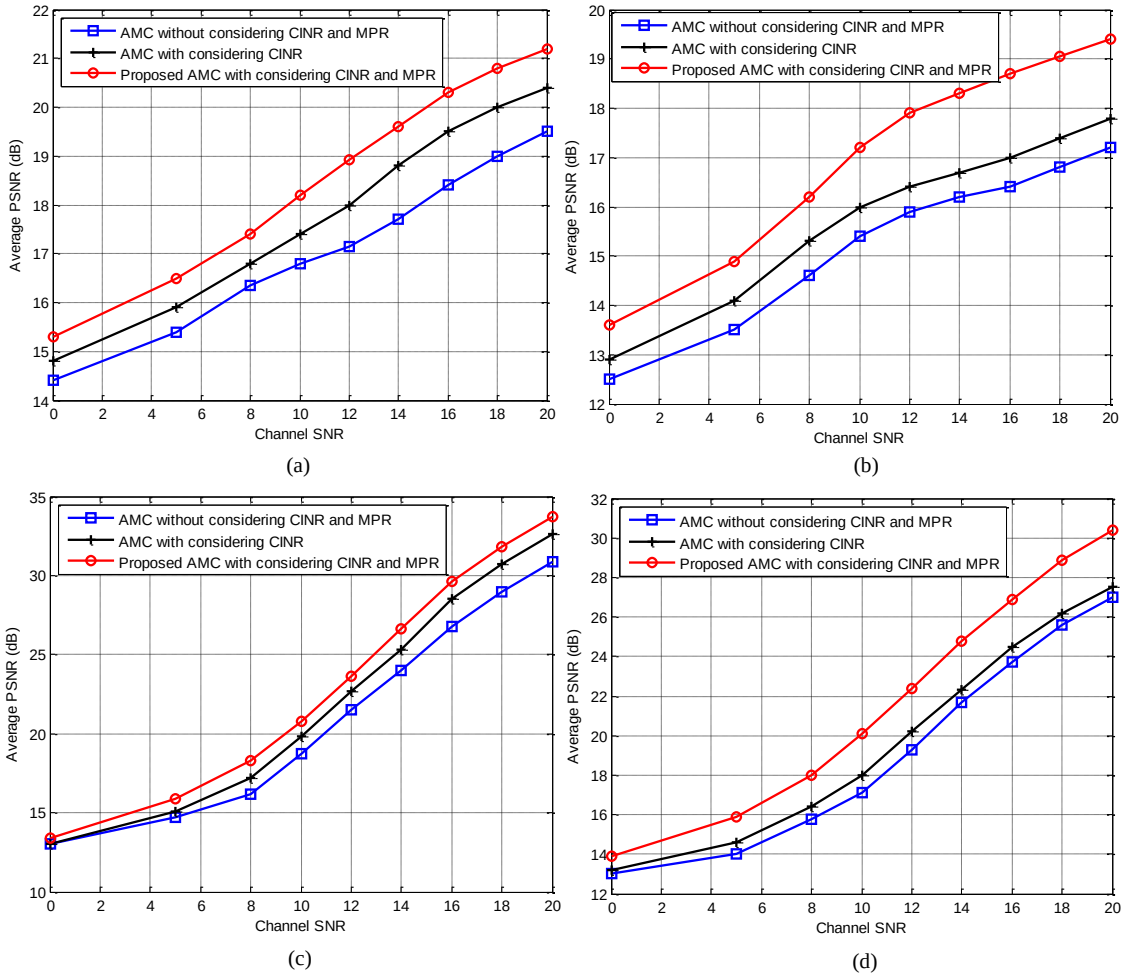



Fig. 6: Simulation result of Stefan and Foreman video transmission over Rayleigh fading channel under two conditions; first condition (I) MPR=3.85, CINR=18, and available bit-rate=6 for CR user in uplink; second condition (II) MPR=1.33, CINR=15, and available bit-rate =5 for CR user in uplink. (a) and (b) are associated with Stefan video under condition (I) and (II) respectively. (c) and (d) are associated with Foreman video under condition (I) and (II) respectively.

Simulation scenario is involved two primary users and one secondary user. All of simulations are performed using Matlab-R2011b software.

## 7. Simulation Resuls

In this study, Quarter pixel Common Inter-mediate Format (QCIF) resolution video sequences Car-phone video, [31], is employed as video sources to evaluate the performance of the proposed scheme, in Uplink, based on the parameters SNR=10.5, MPR=1.5, CINR=23.7. As shown in Fig.5 Frame number 285 of Car-phone video in the secondary user receiver after passing Rayleigh fading channel is as follow; (b) using QAM64 and Code rate = 0.25, (c) using QAM16 and Code rate = 0.5, (d) using QAM and Code rate = 0.67, (e) QAM16 and Code rate= 0.38 and (f) using QAM and Code rate = 0.63 are used for transmission. As shown based on the proposed scheme the best transmission is performed using selecting the QAM16 and Code rate = 0.38.

Next, to evaluate the performance of the proposed scheme and demonstrate the impact of considering CINR and MPR, two sequences Foreman and Stefan [30] are transmitted over Rayleigh fading channel in wide range of channel SNR (0-20 dB). In this simulation we consider three schemes for AMC; in the first scheme we employs AMC scheme which doesn't consider CINR and MPR, second AMC scheme which consider CINR and third AMC scheme which consider both CINR and MPR. For each video we consider two conditions; first condition (I) MPR=3.85, CINR=18, and available bit-rate=6 for CR user in uplink; second condition (II) MPR=1.33, CINR=15, and available bit-rate =5 for CR user in uplink.

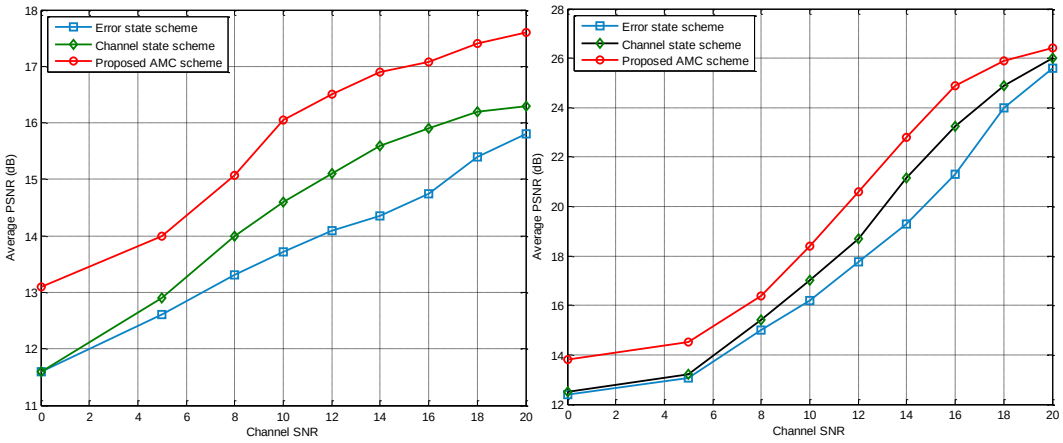



Fig. 7: Simulation results: comparing the proposed scheme with Error state scheme and Channel state scheme, (a) for Stefan sequences (b) for Foreman sequences, over Rayleigh fading channel under condition; MPR=0.25, CINR=10, and available bit-rate=4 for CR user in uplink.

As shown in Fig 6 (a-d) the proposed scheme provides higher PSNR than other two schemes. For example in second condition (II) the AMC scheme which doesn't consider CINR and MPR selects QPSK-0.5, QPSK-0.75, 16QAM-0.5, 16QAM-0.66, 16QAM-0.75, 64QAM-0.5, 64QAM-0.66 and 64QAM-0.75 in SNRs 5, 8, 10, 12, 14, 16, 18, 20 dB respectively [29], but the proposed scheme selects QPSK-0.33, QPSK-0.4, QPSK-0.5, QPSK-0.56, QPSK-0.66, 8QAM-0.66, 8QAM-0.75 and 8QAM-0.75 in SNRs 5, 8, 10, 12, 14, 16, 18, 20 dB respectively. It should be mentioned that the quality of the received video (average PSNR of video) at each Channel SNR is obtained by 15 minute running time.

Next, to compare the proposed scheme with two presented schemes, Channel state scheme and Error state scheme, two sequences Foreman and Stefan [30] are transmitted over Rayleigh fading channel in wide range of channel SNR (0-20 dB). For Channel state scheme, as the best BLER performance is obtained for the lowest target, $BLER_{target} = 10^{-3}$ [30], the target value of BLER is set to $10^{-3}$. For Error state scheme, as the best BLER performance is obtained for Error state memory 3 and 4, in this simulation it is set to 4. In this simulation, Channel state scheme, TBLER algorithm, selects QPSK 1/2 and QPSK 3/4 in SNR ranges 0-15dB and 15-20dB respectively. Error state scheme selects QPSK 1/2, QPSK 3/4, and 16 QAM 1/2 in SNR ranges 0-5dB, 5-15dB and 15-20dB respectively. The proposed scheme selects BPSK 1/4 and QPSK 1/2 in SNR ranges 0-18dB and 18-20dB respectively. As reported in Fig.7, the proposed scheme offers about 1-1.5dB PSNR gain for Stefan and Foreman sequences. This is because of applying accurate AMC model which is affected by considering all effective parameters in modeling procedure, as mentioned before.

## 8. Conclusions

In fast changing environment, a CR system needs a perfect model for AMC. In this paper we have presented a perfect AMC model for OFDM based CR that is compliant with the IEEE.802.16. Presenting a perfect AMC model is advantage of this scheme because of considering all impressive parameters including CINR, available bandwidth, SNR and MPR to select optimum modulation and coding rate. We have modeled the AMC function using ANN because AMC is naturally a non-liner function. In order to achieve more accurate model, GA and PSO have been selected to optimize the function representing relationship between inputs and outputs of ANN, i.e., AMC model. Simulation results prove that the proposed AMC scheme presents perfect and powerful decision to select optimum modulation and coding rate and consequently provides higher quality for delivered video. Also it was shown GA is more powerful optimizer algorithm than PSO.

## 9. Acknowledgment

This work has been supported by the Research Institute for Information and Communication Technology (ITRC), Tehran, Iran under Grant, T/19262/500.

## References


[1] Steinbach, Eckehard, Niko Farber, and Bernd Girod. "Standard compatible extension of H. 263 for robust video transmission in mobile environments." Circuits and Systems for Video Technology, IEEE Transactions on 7.6 (1997): 872-881. 872881, Dec. 1997.

[2] Hsu, Chi-Yuan, Antonio Ortega, and Masoud Khansari. "Rate control for robust video transmission over burst-error wireless channels." Selected Areas in Communications, IEEE Journal on (1999): 756-773.

[3] Zhai, Fan, and Aggelos Katsaggelos. "Joint source-channel video transmission." Synthesis Lectures on Image, Video, and Multimedia Processing (2007)

[4] Mitola III, Joseph, and Gerald Q. Maguire Jr. "Cognitive radio: making software radios more personal." Personal Communications, IEEE 6.4 (1999): 13-18.

[5] Arslan, Hüseyin, ed. Cognitive radio, software defined radio, and adaptive wireless systems. Springer London, Limited, 2007.

[6] H. Shatila, M. Khedr, J. H. Reed "Cognitive Radio, Software Defined Radio, and Adaptive Wireless Radio," International Symposium on Performance Evaluation of Computer and Telecommunication Systems, 2010

[7] Yang, James, Noel Tin, and Amir K. Khandani. "Adaptive modulation and coding in 3G wireless systems." Vehicular Technology Conference, 2002. Proceedings IEEE, 2002.

[8] Ahn, Seok-Ki, and Kyeongcheol Yang. "Adaptive modulation and coding schemes based on LDPC codes with irregular modulation." Communications, IEEE Transactions on 58.9 (2010): 2465-2470.

[9] Liang, Yu-Cheng, Ching-Chun Chou, and Hung-Yu Wei. "Modeling and analysis of applying adaptive modulation coding in wireless multicast and broadcast systems." Wireless Networks (2011): 1373-1386.

[10] El Shabrawy, Tallal Osama, and Sherif Hassan Abdel Wahed. "Adaptive Power Saving Receiver for DVB-H Exploiting Adaptive Modulation and Coding." IJCNS 3.2 (2010): 152-166.

[11] Jang, Edward W., Chan-Soo Hwang, and John M. Cioffi. "Flexible Adaptive-Modulation-and-Coding tables for a wireless network." Communications, 2008. ICC'08. IEEE International Conference on. IEEE, 2008.

[12] Kang, Chung G., Si H. Park, and Jin W. Kim. "Design of adaptive modulation and coding scheme for truncated hybrid arq." Wireless personal communications 53.2 (2010): 269-280.

[13] Armanious, Ehab, David D. Falconer, and Halim Yanikomeroglu. "Adaptive modulation, adaptive coding, and power control for fixed cellular broadband wireless systems: some new insights." Wireless Communications and Networking,. IEEE, 2003.

[14] She, James, Jingqing Mei, James Ho, Pin-Han Ho, and Hong Ji. "Layered adaptive modulation and coding for 4G wireless networks." In Global Telecommunications Conference, 2010 IEEE, pp. 1-6., 2010..

[15] Goldsmith, Andrea J., and S-G. Chua. "Adaptive coded modulation for fading channels." Communications, IEEE Transactions on (1998): 595-602.

[16] Goldsmith, Andrea J., and Pravin P. Varaiya. "Capacity of fading channels with channel side information." Information Theory, IEEE Transactions on 43.6 (1997): 1986-1992.

[17] Goldsmith, Andrea J., and Soon-Ghee Chua. "Variable-rate variable-power MQAM for fading channels." Communications, IEEE Transactions on 45.10 (1997): 1218-1230.

[18] S. M. Alamouti and S. Kallel "Adaptive trellis-coded multiple-phased shift keying for Rayleigh fading channels," IEEE Trans. Commun, vol. 42, pp. 2305-2314, 1994.

[19] M. Z. Bocu, J. P. Coon, C. N. Canagarajah, J P. McGeehan, S. M. D. Armour and A. Doufexi "resource allocation for OFDMA-Based cognitive radio networks with application to H.264 scalable video transmission," EURASIP Journal on wireless communications and networking , 245-673, 2011.

[20] Tsai, Ming-Fong, et al. "MAC-level forward error correction mechanism for minimum error recovery overhead and retransmission." Mathematical and computer modelling 53.11 (2011): 2067-2077.

[21] Tsai, Ming-Fong, et al. "Concurrent multipath transmission combining forward error correction and path interleaving for video streaming." Computer Communications 34.9 (2011): 1125-1136.

[22] Richardson, Iain E. H. 264 and MPEG-4 video compression: video coding for next-generation multimedia. Wiley. com, 2004.

[23] Sullivan, Gary J., Pankaj N. Topiwala, and Ajay Luthra. "The H. 264/AVC advanced video coding standard." Optical Science and Technology, the SPIE 49th Annual Meeting. International Society for Optics and Photonics, 2004.

[24] Beasley, David, R. R. Martin, and D. R. Bull. "An overview of genetic algorithms: Part 1. Fundamentals." University computing (1993): 58-58..

[25] M. T. Hagan "Neural Network Design, Oklahoma State University," 2002.

[26] J. Kennedy and R. C. Eberhart, "Particle Swarm Optimization," In Proc. IEEE International Conference on Neural Networks IV, 1942-1948, 1995.

[27] Clerc, Maurice, and James Kennedy. "The particle swarm-explosion, stability, and convergence in a multidimensional complex space." Evolutionary Computation, IEEE Transactions on 6.1, 58-73., 2002

[28] Zahiri, Seyed-Hamid. Swarm intelligence and fuzzy systems. Nova Science Publishers, Inc., 2010..

[29] IEEE Standard for Local and metropolitan area networks Part 16, IEEE Standard 802.16, 10016-5997, USA, 2009.

[30] Fantacci, Romano, et al. "Adaptive modulation and coding techniques for OFDMA systems." Wireless Communications, IEEE Transactions on (2009): 4876-4883.

[31] [Online] Available: http://media.xiph.org/video/derf/

[32] Reza Bahreini Behzadi and Ali Asghar Aslaminejad, 2010. A Comparison of Neural Network and Nonlinear Regression Predictions of Sheep Growth. Journal of Animal and Veterinary Advances, 9: 2128-2131.

[33] Guidelines for evaluation of radio transmission technologies for IMT-2000, ITU-R Recommendation M.1225, Feb. 1997.

**Hasan Farsi** received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Tehran, Iran, in 1992 and 1995, respectively. Since 2000, he started his Ph.D in the Centre of Communications Systems Research (CCSR), University of Surrey, Guildford, UK, and received the Ph.D degree in 2004. He is interested in speech, image and video processing on wireless communications. Now, he works as associate professor in communication engineering in department of Electrical and Computer Eng., university of Birjand, Birjand, IRAN. His Email is: hfarsi@birjand.ac.ir

**Farid Jafarian** received his B.S. (with highest honors) degree from Isfahan's Sepahan Institute of Science and Technology, Isfahan, Iran and M.S. degree from University of Birjand, Birjand, Iran, in 2008 and 2012 respectively, all in Electrical Engineering. He was with the Research Institute for Information and Communication Technology (ICT), Tehran, Iran. He is currently senior lecturer in Bonyan University. His research interest and contributions are in the areas of communication theory, information theory and their applications to wireless communications (e.g., multiuser MIMO, cognitive radio and cooperative communication), joint source and channel coding (JSCC) and communications signal processing with a focus on video and speech coding. He is also interested in applications of information theory in Bioinformatics with a focus on genome modeling.

# EBG Structures Properties and their Application to Improve Radiation of a Low Profile Antenna

Masoumeh Rezaei Abkenar
Department of Electrical and Computer Engineering, Semnan University
m_rezaie83@yahoo.com
Pejman Rezaei*
Department of Electrical and Computer Engineering, Semnan University
prezaei@semnan.ac.ir

Received: 24/Nov/2012 Accepted: 11/Dec/2013

**Abstract**
In this paper we have studied the characteristics of mushroom-like Electromagnetic Band Gap (EBG) structure and performance of a low profile antenna over it. Afterward, a novel EBG surface is presented by some modifications in mushroom-like EBG structure. This structure, which has more compact electrical dimensions, is analyzed and its electromagnetic properties are derived. Results show that resonant frequency of this novel structure is about 15.3% lower than the basic structure with the same size. Moreover, the novel EBG structure has been used as the ground plane of antenna. Its application has improved radiation of a low profile dipole antenna. The antenna performance over the new EBG ground plane is compared with the conventional mushroom-like EBG structure. Simulation results show that using this slot loaded EBG surface, results in 13.68dB improvement in antenna return loss, in comparison with conventional mushroom-like EBG, and 33.87dB improvement in comparison with metal ground plane. Besides, results show that, EBG ground planes have increased the input match frequency bandwidth of antenna.

**Keywords:** Electromagnetic Band Gap (EBG), Low Profile Antenna, Slot Loaded EBG Surface, Bandwidth, Dipole Antenna.

## 1. Introduction

In the last two decays, artificial periodic structures have been used in a wide range of engineering applications. Electromagnetic Band Gap (EBG) structures are a group of these artificial periodic structures, and recently their application in different antennas has attracted much research interests in electromagnetic applications. EBG structures are a novel class of artificially fabricated structures, and can control the propagation of electromagnetic waves inside themselves. These structures have two important and special electromagnetic properties. The first one is suppression of surface waves in a specific frequency band, which called the band gap. The other one is phase response to the plane wave illumination; these structures have a reflection phase that changes vs. frequency from 180º to -180º [1]. Besides, these structures possess some other exciting features, like high impedance in their performance band. According to these properties, a wide range of applications have been reported, such as TEM waveguides, different microwave filters and low profile wire antennas [2-5].

Mushroom-like EBG structure initially designed by D. Sievenpiper [6] is a popular structure that exhibits compactness and simple implementation features compared to other EBG structures. This conventional mushroom-like EBG can be used in different antenna designs to suppress surface waves. But in some practical applications smaller cell size is needed.

In this paper, at first we have studied the features of a mushroom-like EBG structure. Then, a novel structure is designed by inserting some slots in the patches of mushroom-like EBG cells. These slots significantly enlarge capacitance of equivalent LC circuit and so result in a more compact structure. Electromagnetic properties of the new slot loaded EBG is derived and compared with conventional mushroom-like EBG. Additionally, to study the effect of the EBG structures, performance of a low profile antenna over these structures has been observed. We have utilized both EBG structures as ground planes to improve the radiation efficiency of a dipole antenna near the ground plane. Also, performance of antenna has been compared with an antenna over Perfect Electric Conductor (PEC) ground plane.

A considerable improvement in antenna performance has been observed.

The structure was analyzed in our previous works [7-10] by Ansoft HFSS, which is a commercially available simulation tool based on finite element method [11]. In this paper the dispersion diagram of the structure is also derived by CST Microwave Studio [12], and the result are in agreement with previous results.

* Corresponding Author

## 2. Surface Waves

In this section we want to study surface waves and the conditions in which they can occur. Surface waves are the waves which can exist on the interface between any two dissimilar materials, like metal and free space. They are strongly bounded to the surface, and their fields exponentially decay along normal direction to the surface [13]. Surface waves are an important issue for many antennas, since these waves propagate along the ground plane instead of radiation into free space, so reduce the antenna efficiency and gain.

To derive the characteristics of surface waves on the interface between a material and free space, assume the surface is in the YZ plane, free space extending in the +X direction, and the other material in the –X direction. This configuration is shown in Figure 1.

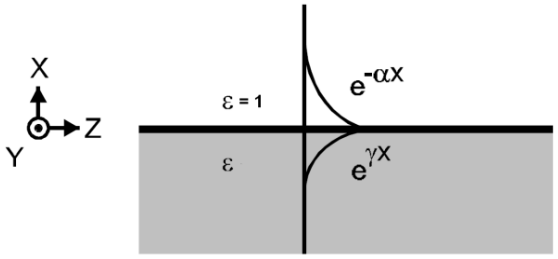



Fig. 1 A surface wave on the interface of a material and free space.

The surface wave decays in the +X direction with decay constant α, and in the –X direction with decay constant γ. By combining electromagnetic fields in two materials according to Maxwell's equations, decay constants α and γ for a TM surface wave are derived as (1) and (2) [14,15]:

$$\alpha = \sqrt{-1/(1+\varepsilon)}\ \omega/c \tag{1}$$

$$\gamma = \sqrt{-\varepsilon^2/(1+\varepsilon)}\ \omega/c \tag{2}$$

For a positive ε, decay constants are imaginary, and the waves do not decay as the distance from the surface increases. Thus, these waves are plane waves that propagate through the dielectric interface. On the other hand, when ε is less than –1, or when it is imaginary, the solution shows a wave that is bound to the surface. Therefore, TM mode surface waves can exist only on the metals, or other materials with non-positive dielectric constants. By exchanging the electric and magnetic fields and substituting ε with μ, according to the principle of duality, above expressions can be applied to the TE mode [16].

From the other point of view we can consider the surface impedance of surface, which is defined as the ratio of the electric field over the magnetic field. For the above surface in the YZ plane, required surface impedance for TM surface waves is obtained as below by considering electric and magnetic fields on the surface [14,17]:

$$Z_s(TM) = \frac{E_z}{H_y} = \frac{j\alpha}{\omega\varepsilon} \tag{3}$$

So TM waves only can occur on a surface with positive reactance that means inductive surface impedance. For a TE wave the surface impedance is equal to the following expression:

$$Z_s(TE) = \frac{E_z}{H_y} = -\frac{j\omega\mu}{\alpha} \tag{4}$$

Thus, a negative reactance is necessary for TE surface waves, it means capacitive surface impedance.

## 3. Reflection Phase

The Reflection phase is an important property of EBG structures, which is determined as the phase of reflected electric field at the reflecting surface. In EBG structures it varies with frequency from 180º to -180º. So in a specific frequency band, when the reflection phase is around 0º, they can be used as proper ground planes, like Perfect Magnetic Conductors (PMC).

For a surface in the YZ plane, the surface impedance seen by an incident wave in the X direction is equal to:

$$Z_s = E_z / H_y \tag{5}$$

For a high impedance surface, the ratio in Equation (5) is too high, so the electric field has a non-zero value, while the magnetic field is zero. In the other word, this surface is called a magnetic conductor, because of its zero tangential magnetic fields at the surface.

The reflection phase is the phase difference between the backward and forward waves which formed a standing wave on an arbitrary surface. The electromagnetic fields on the surface are expressed as (6) and (7). Besides, the boundary condition at the surface is given by the surface impedance as (8) [18]:

$$E(x) = E_f\ e^{-jkx} + E_b\ e^{jkx} \tag{6}$$

$$H(x) = H_f\ e^{-jkx} + H_b\ e^{jkx} \tag{7}$$

$$\frac{E_{total}(x=0)}{H_{total}(x=0)} = Z_s \tag{8}$$

Moreover, the electric and magnetic fields of each wave are related by means of the impedance of free space as (9).

$$\left|\frac{E_f(x)}{H_f(x)}\right| = \left|\frac{E_b(x)}{H_b(x)}\right| = \sqrt{\frac{\mu_0}{\varepsilon_0}} = \eta \tag{9}$$

And the reflection phase is equal to:

$$\Phi = \mathrm{Im}\left\{\ln\left(\frac{E_b}{E_f}\right)\right\} \tag{10}$$

Combining Equation (8), (9) and (10) gives the reflection phase of a surface with impedance Zs:

$$\Phi = \mathrm{Im}\left\{\ln\left(\frac{Z_s-\eta}{Z_s+\eta}\right)\right\} \tag{11}$$

As a result when Zs has a low value, like a PEC surface, the reflection phase will be ± π, and when it is very high, like an EBG surface, the reflection phase will be zero.

## 4. EBG Structures Design and Characterization

### 4.1 Mushroom-Like EBG

As previously mentioned, mushroom-like EBG structure is one of the basic and most common EBG structures. This structure is shown in Figure 2. It consists of a flat metal sheet that is covered with an array of metal protrusions on a dielectric substrate which are connected to the lower conducting surface by metal vias [6]. The parameters of the EBG structure are labeled as patch width w, gap width g, substrate thickness h, dielectric constant εr, and vias radius r. When the periodicity of structure, which is equal to w+g, is small compared to the operating wavelength, the operation mechanism of structure can be explained using an effective medium model with equivalent lumped LC elements, as explained in [18]. The capacitor C, results from the gap effect between the patches and the inductor L, due to the current flowing along adjacent patches. Thus, the surface impedance and central frequency of band gap are estimated like a parallel resonant circuit:

$$Z = \frac{j\omega L}{1-\omega^2 LC} \quad (12)$$

$$\omega_0 = 1/\sqrt{LC} \quad (13)$$

It is clear that, at low frequencies surface impedance of the structure is inductive, so it supports TM surface waves. Inversely, it is capacitive at high frequencies, and supports TE surface waves. Around the LC resonance frequency, the impedance is very high. In this frequency band the structure suppresses propagation of both TE and TM modes of surface waves. Also, it reflects incident electromagnetic waves without phase reversal that occurs on a PEC. This frequency band is called the band gap.

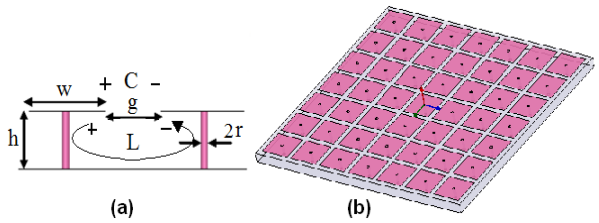



Fig. 2 Mushroom-like EBG, a) unit cell, b) 3D view.

### 4.2 Slot Loaded EBG

Since Slot loaded EBG is a new type of EBG structures which are designed by cutting some slots into the metal patches of conventional mushroom-like EBG. These slots change the current flow on the patches which caused to a longer current path. They also create extra capacitance between the slot edges. So the values of L and C in (1) are increased and result in a lower frequency band gap and finally a more compact structure [19-23].

In this paper we have designed a novel slot loaded EBG structure by cutting a pair of I-like, X-oriented slots into the patch of a mushroom-like EBG cell. The structure is shown in Figure 3. Dimensions of the basic mushroom-like cell are designed as: w=3mm, g=0.5mm, h=1mm and r=0.125mm. A dielectric layer with εr=2.33 is used as substrate. Lengths of slots are optimized to obtain the most compact structure. Finally, dimensions of slots are designed as: L1=1.5mm, L2=1mm, L3=0.5mm and L4=0.25mm.

The initial inductance and capacitance of the mushroom-like EBG structure are [18]:

$$L = \mu_0 \mu_r h \quad (14)$$

$$C = \frac{W\varepsilon_0(1+\varepsilon_r)}{\pi}\cosh^{-1}\left(\frac{W+g}{g}\right) \quad (15)$$

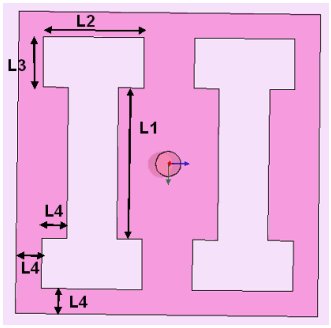



Fig. 3 Unit cell of the designed EBG cell.

The equivalent inductance and capacitance of the new structure are equal to the above initial inductance and capacitance, in addition to the new inductance and capacitance which are created by slots. So by inserting the slots, the initial value of L and C remain unchanged, while the equivalent L and C will increase, and result in a lower resonant frequency. Thus, the wavelength will increase and the electrical dimensions of the EBG cell, which mean the dimensions in comparison to the wavelength, will decrease. Thus, we will achieve to a more compact EBG cell. In next section we will study this structure's properties, and it will be compared with the basic structure.

### 4.3 Characterization of EBG Structures

#### 4.3.1 Reflection Phase Diagram

As it is mentioned in previous sections, reflection phase is one of the most important characteristics of EBG structures. To extract the reflection phase diagram in HFSS, a unit cell is modeled and periodic boundary conditions are applied on side walls. This structure is shown in Figure 4. The cell is excited by a plane wave in different frequencies and phase of the reflected wave at evaluation plane is calculated as [24]:

$$\Phi_{EBG} = \frac{\int_S \text{Phase}(E_{scattered})ds}{\int_S \hat{s}ds} \quad (16)$$

In the above expression, S is the evaluation plane.

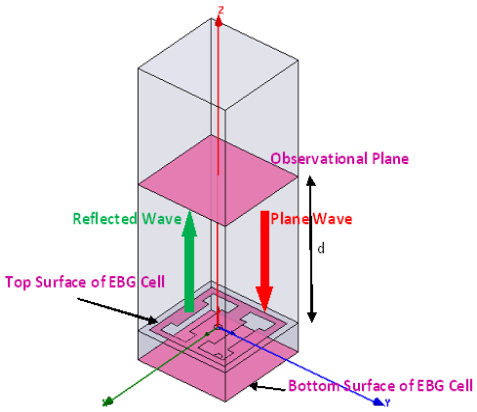



Fig. 4 Simulated cell to extract reflection phase.

The reflection phase diagrams of a normally incident plane wave on the conventional mushroom-like EBG and the new slot loaded EBG are obtained by this method and shown in Figure 5.

In antenna applications, it has been shown practically that the desired band for antenna radiation on an EBG plane is close to the frequency region where the EBG surface has a reflection phase in the range of 90º±45º [25]. The 90º±45º criterion is also compatible to PEC and PMC planes. PEC surface has 180º reflection phases for a horizontally positioned dipole antenna, and reverse image current decreases the antenna radiation performance, while a PMC surface has 0º reflection phase. Hence, when the EBG ground plane exhibits a reflection phase in the middle of this region, a good return loss is expected for the dipole antenna.

According to Figure 5, 90º±45º frequency band and in-phase (zero degree) reflection frequency of mushroom-like and slot loaded EBG structures are shown in Table 1. So the in-phase reflection frequency of novel slot loaded EBG is 15.3% lower than mushroom-like EBG of the same size, and decreases from 17GHz to 14.4GHz, since the slots affect the electric currents flowing along the patch and result in a longer current path [18].

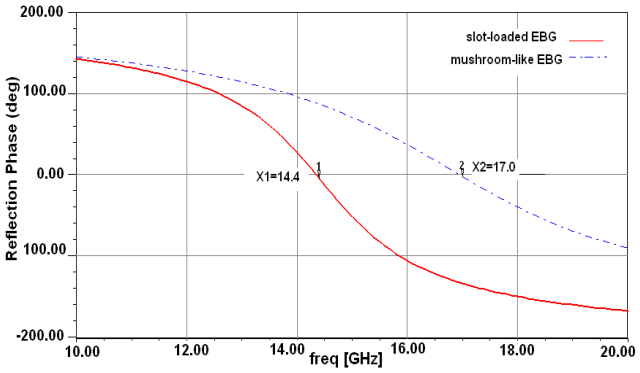



Fig. 5 Reflection phase of a normally incident plane wave with Ex polarization on the two types of EBG structures.

Table 1: 90º±45º frequency band and in-phase reflection frequency of mushroom-like and slot loaded EBG structures

| Structure Types | Reflection Phase Specifications [GHz] | | |
|---|---|---|---|
| | fl | fh | f0 |
| Mushroom-like EBG | 11.3 | 15.8 | 17 |
| Slot loaded EBG | 10.8 | 13.8 | 14.4 |

According to Table 1 the period of mushroom-like and slot loaded EBG structures at their in-phase reflection frequency, f0, and central frequency of 90º±45º region, fc, is calculated and shown in Table 2. It represents that the slot loaded EBG surface has smaller cell size at different operation frequencies.

Table 2: Period of EBG structures

| Structure Types | Period of Structure [λ] | |
|---|---|---|
| | f0 | fc |
| Mushroom-like EBG | 0.198 λ0 | 0.158 λc |
| Slot loaded EBG | 0.168 λ0 | 0.144 λc |

On the other hand, the results show that if we increase the patch size of mushroom-like EBG cell to 3.85mm, the structure will operate at the same frequency band to the slot loaded EBG surface. In this case the period of structure is P=w+g=4.35mm, so the new slot loaded EBG cell size is reduced by 20%. This structure's reflection phase is shown in Figure 6.

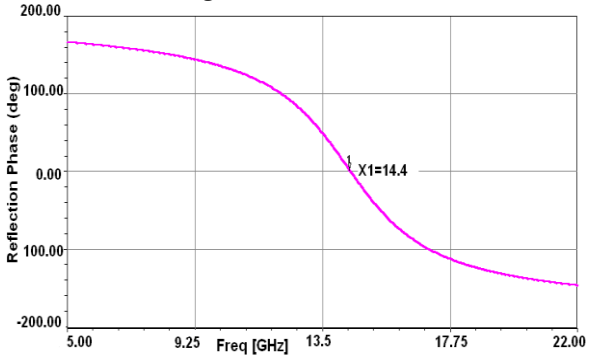



Fig. 6 Reflection phase of a Mushroom-like EBG cell with w=3.85mm.

### 4.3.2 Dispersion Diagram Method

The Dispersion diagram or β-f, curve is the other important characteristic of EBG structures, and can be calculated from the unit cell and applying a periodic boundary condition with appropriate phase shifts on the sides. According to Floquet theorem and its expansion by Bloch, dispersion curve is periodic. Therefore, we only need to plot the dispersion relation within one single period, which is known as the Brillouin zone. The smallest region within the Brillouin zone for which the directions are not related by symmetry is called the irreducible Brillouin zone. The irreducible Brillouin zone of our structure is a triangular wedges with 1/8 the area of the full Brillouin zone defined by Γ, X and M points [26]. The simulated cell by CST Microwave Studio [12], is depicted in Figure 7.

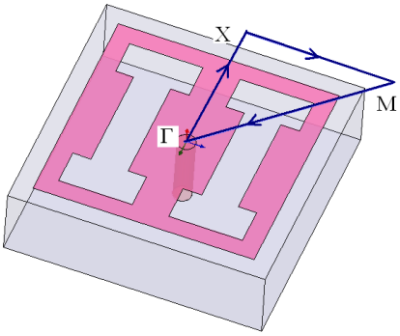



Fig. 7 The simulated cell to extract dispersion diagram.

Two dimensional Eigen mode solutions for Maxwell's equations are solved for the Brillouin zone. As it can be seen in Figure 8, the band gap of dispersion curve for the mushroom-like EBG is between 11.48-16.14GHz, while for the slot-loaded EBG it is about 10.81-15.30GHz frequency band. Thus, the novel slot-loaded EBG has a band gap in a lower frequency band, in comparison to the mushroom-like EBG.

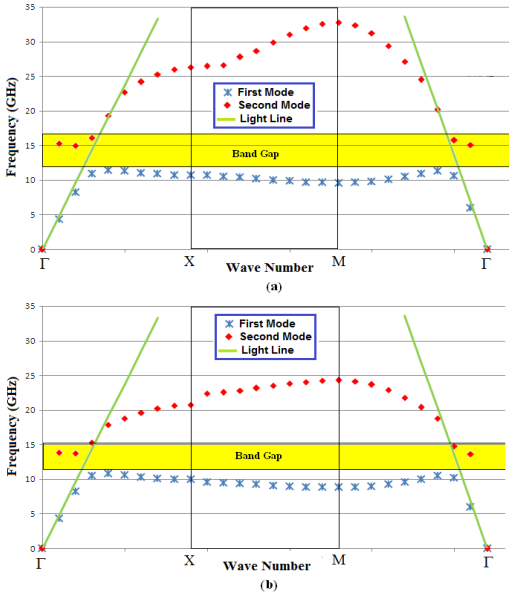



Fig. 8 Dispersion diagram, a) mushroom-like EBG, b) slot loaded EBG.

### 4.3.3 Direct Transmision Method

The other method to determine the band gap of EBG structure is direct transmission. In this method wave transmission through EBG structure is modeled as scattering parameter S21. As it is shown in Figure 9, a part of structure which is repeated in different directions of the lattice is considered as a TEM waveguide with two pairs of parallel PEC and PMC sides [27].

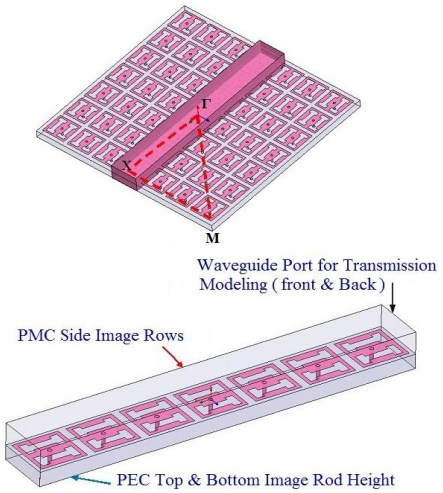



Fig. 9 The simulated TEM waveguide in direct transmission method.

We excite it with two ports on remained sides, and the transmission coefficient between these two ports can show the band gap. But this method is appropriate just for symmetric structures and along the propagation direction. In this paper we have used a 7×7 lattice of mushroom-like and the slot loaded cells near antenna. So we have a 1×7 lattice as a TEM waveguide. The transmission diagram of these structures is shown in Figure 10. The desired frequency band usually is considered as the region where S21 has the value less than -20dB.

According to Figure 10, the band gap of mushroom-like EBG is 10.39-15.73GHz, while it is 10.06-14.07GHz for the slot loaded EBG. Hence, it is clear that the band gap region of the novel structure is decreased based on direct transmission method too.

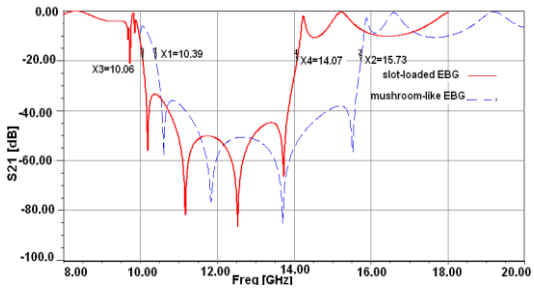



Fig. 10 Comparison of S21 for mushroom-like and slot loaded EBG.

In this section, the EBG structure is analyzed with three methods. Because of the different nature of these methods, the results are not necessarily the same. In fact each of them is measuring a different property of the structure, and depending on the application one of them will be more useful. For example when the structure is used as the ground plane for a wire antenna, we want to have in phase reflection, so reflection phase method is more applicable. Moreover, we can call the overlap of these three as the bandwidth of the structure, to be sure that we have selected the right working frequency for the antenna.

## 5. Low Profile Dipole Antenna with EBG Ground Plane

In many communication devices, it is more desirable to have a low profile antenna. In these antennas the overall height is usually less than one tenth of the operating wavelength [1]. Besides, in many antennas a metal plane is used as a reflector or ground plane [16]. This ground plane redirects one-half of the radiation into the opposite direction, improving the antenna gain by 3dB, and shielding objects on the other side [6], but it causes a limitation for the antenna's height. The antenna can't be too close to the metal plane because of the coupling effect. So a fundamental challenge is the coupling effect of the ground plane.

If we place a vertical antenna over a PEC surface as the ground plane, the electric current will be vertical to the plane, so the image current will be in the same direction and will reinforce the radiation of the original current. Thus, this antenna has good radiation efficiency, but it suffers from high height of antenna, due to the vertical placement. To realize a low profile configuration, it is better to put the antenna horizontally over the ground plane, but in this case, the problem will be the poor radiation efficiency because of 180º reflection phase of PEC. As it is mentioned in previous sections, the EBG surface has a variable reflection phase, and is capable of providing a constructive image current within a certain frequency band. So it can result in good radiation efficiency.

As Figure 11 shows, we have placed a dipole antenna horizontally over an EBG ground plane with 7×7 array of cells. The radius of the dipole is 0.125mm and the height of the dipole is 0.5mm, so the overall height of antenna structure is 1.5mm. In order to obtain a resonant condition for a half-wave dipole at 12GHz, the physical length must be somewhat shorter than the free space half-wavelength [28]. To find the optimum length of dipole, it has been changed and each time return loss of the antenna is obtained. Results show that the best return loss is achieved by a dipole with length of 12mm. Return loss of this dipole antenna over mushroom-like and slot loaded EBG ground planes is shown in Figure 12.

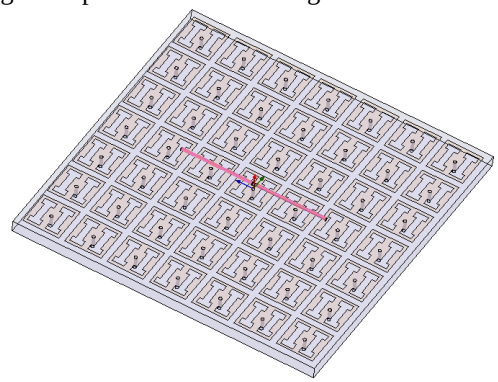


Fig. 11 Dipole antenna over EBG ground plane.

To study the effect of EBG structure first we place the dipole over a PEC ground plane of the same size as the EBG ones. The least return loss is -2.48dB at 11.58GHz, as it is shown in Figure 12. Then, we have used the mushroom-like EBG surface. In this case, the antenna has a return loss of -22.67dB at 12.64GHz. Finally, by replacing the mushroom-like EBG with the slot loaded structure, the antenna has showed the resonant frequency of 11.88GHz, and in this frequency return loss is-36.35dB.

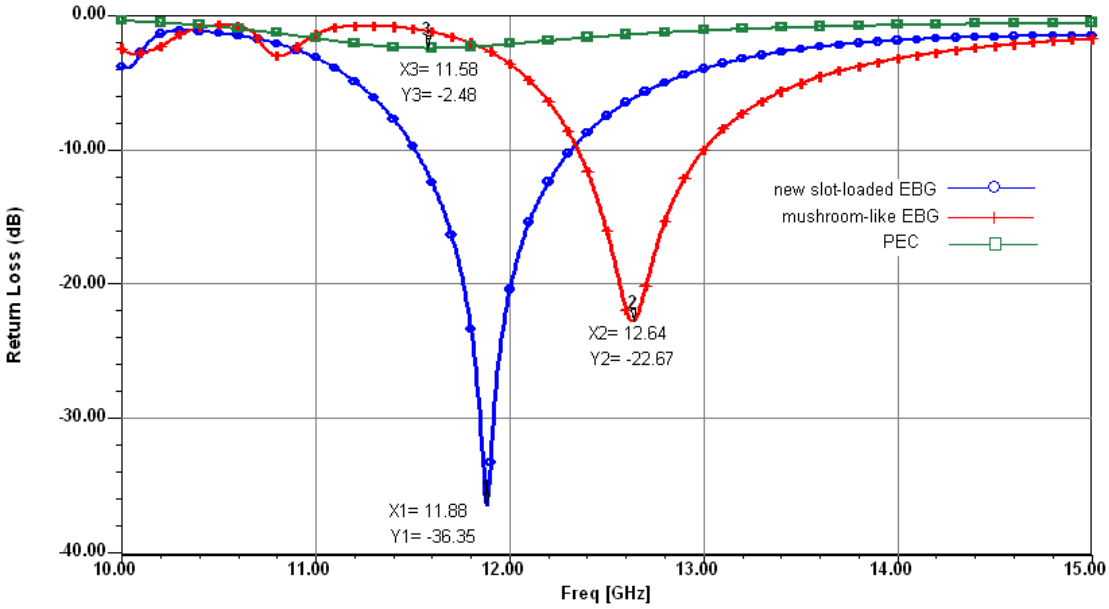



Fig. 12 Comparison of antenna return loss with mushroom-like EBG (P=w+g=3.5mm), slot-loaded EBG, and PEC ground planes.

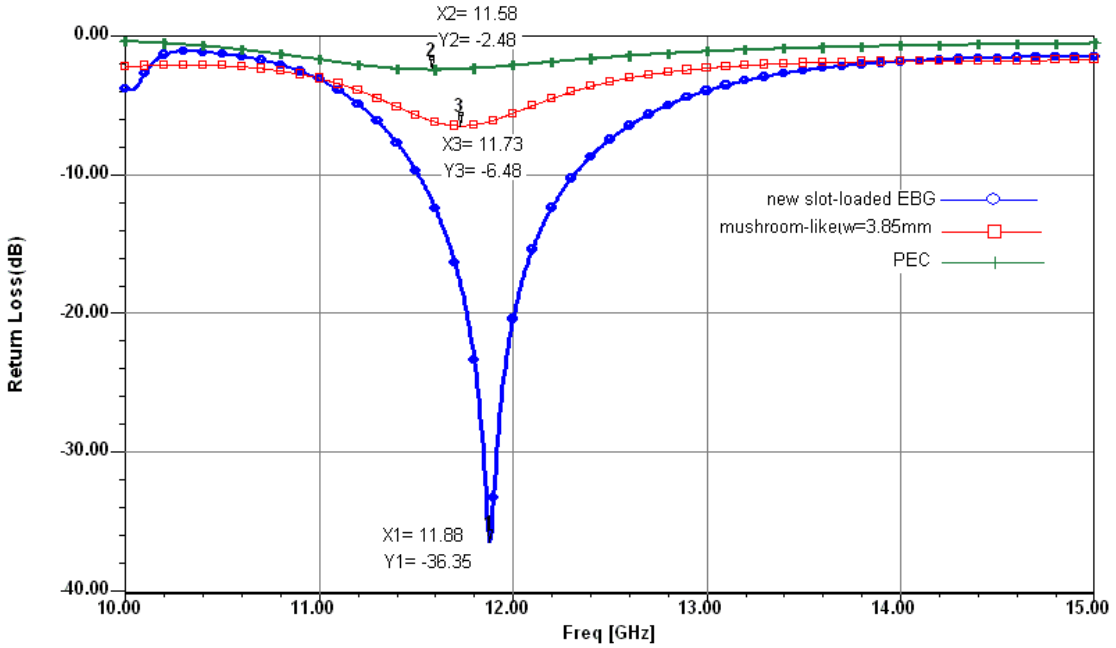



Fig. 13 Comparison of antenna return loss with mushroom-like EBG (P=w+g=4.35mm), slot-loaded EBG and PEC ground planes.

It is clear that using EBG ground planes have improved the return loss of antenna, and this improvement is more considerable for the slot loaded surface. By using the slot loaded EBG, we have 33.87dB improvement in antenna return loss in comparison to the PEC ground plane, and 13.68dB increase in comparison to the conventional mushroom-like structure.

The input-match frequency band of antenna is defined as the region where the antenna has a good return loss, usually less than -10dB. Regarding Figure 12, this bandwidth which can't be achieved by PEC ground plane, is 5.14% for the mushroom-like EBG and 6.82% for the slot loaded EBG.

As it is discussed in the previous sections, a mushroom-like EBG with the period of 4.35mm has the same frequency response as the novel slot loaded EBG structure. We have compared them as the ground plane of dipole antenna, too. Because of greater patch size of this mushroom-like cell, we used a 5×5 array to preserve the overall size of ground plane. The antenna return loss is depicted in Figure 13, and it is compared with the slot loaded and PEC ground planes. By this structure, the resonant frequency of antennas is almost the same, but the antenna has the minimum return loss of -6.48dB at 11.73GHz, and the input match frequency bandwidth is not accessible. Thus, it is not a good substitute for the novel slot loaded EBG structure, since the performance of the slot loaded surface is better. The results of return loss curves for different types of ground planes are summarized in Table 3.

Table 3: Comparison of performance of four different ground planes

| Ground Plane Type | Resonant frequency (GHz) | Minimum return loss (dB) | Bandwidth |
|---|---|---|---|
| Novel slot loaded (P=3.5m) | 11.88 | -36.35 | 6.82% |
| Mushroom-like (P=3.5mm) | 12.64 | -22.67 | 5.14% |
| Mushroom-like (P=4.35mm) | 11.73 | -6.48 | 0% |
| PEC (25×25mm) | 11.58 | -2.48 | 0% |

Poor return loss of the PEC ground plane is due to 180º reflection phase and reverse image current, while using EBG ground plane, radiation of antenna has improved because of desirable reflection phase and surface wave frequency band gap. Figure 14 shows surface current density at resonant frequency. In EBG structure, surface current has been reduced because of its frequency band gap. Moreover, According to lower frequency band of the slot loaded EBG surface, it has a smaller cell size in comparison to mushroom-like EBG. The advantage of compactness makes this new EBG structure a promising candidate in practical applications.

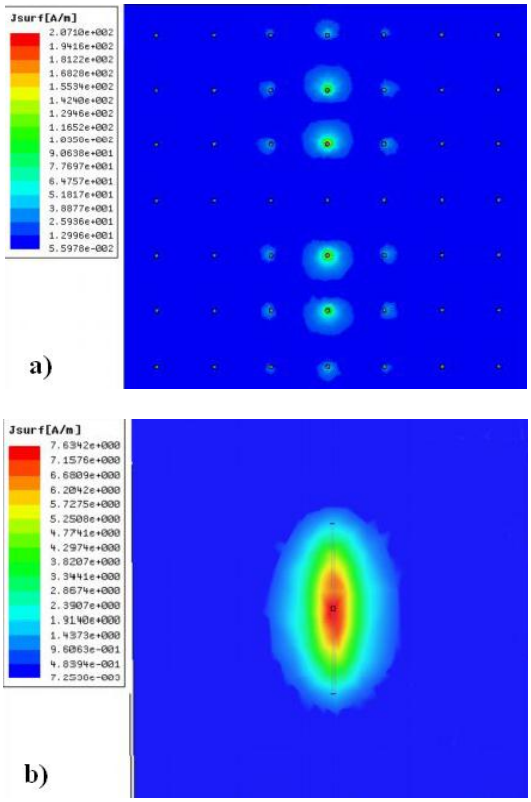



Fig. 14 Surface current density: a) EBG, and b) PEC ground planes.

## 6. Conclusions

In this paper, we have studied the properties of EBG structures, and have presented a novel slot loaded EBG. Its reflection phase and transmission characteristics are derived and compared with conventional mushroom-like EBG surface. Simulation results show that in-phase reflection frequency of this slot loaded EBG is 15.3% lower than mushroom-like EBG structure of the same size, and decreases from 17GHz to 14.4GHz. So the designed EBG is a compact structure and is a proper candidate in practical applications. Besides, we have used EBG structures to design a low profile antenna. These structures have been used as the ground plane to improve radiation of a low profile dipole antenna. Although both mushroom-like and slot loaded structures improve the return loss, the novel slot loaded EBG has better results. Using this slot loaded EBG surface, the antenna return loss decreases more than 13.68dB in comparison with conventional mushroom-like EBG of the same size, and 33.87dB in comparison with PEC ground plane. Moreover, the input-match frequency band of antenna is improved by 1.68% by replacing the mushroom-like EBG ground plane of the same size with the novel slot loaded EBG structure.

### Acknowledgments

The authors would like to acknowledge the assistance and financial support provided by the Semnan University.

## References

[1] F. Yang, and Y. Rahmat-samii, Electromagnetic band gap structures in antenna engineering, Cambridge University Press, 2008.

[2] D.Y. Shchegolkov, C.E. Heath, and E.I. Simakov, "Low loss metal diplexer and combiner based on a photonic band gap channel-drop filter at 109 GHz," Progress In Electromagnetics Research, PIER 111, 2011, pp. 197-212.

[3] T.-N. Chang, and J.-H. Jiang, "Enhance gain and bandwidth of circularly polarized microstrip patch antenna using gap-coupled method," Progress In Electromagnetics Research, PIER 96, 2009, pp. 127-139.

[4] F. Yang and Y. Rahmat-Samii, "Wire antenna on an EBG ground plane vs. patch antenna: A comparative study on low profile antennas," URSI Electromagnetic Theory Symposium, Ottawa, Canada, July 2007.

[5] D. Zarifi, H. Oraizi, and M. Soleimani, "Improved Performance of Circularly Polarized Antenna Using Semi-Planar Chiral Metamaterial Covers," Progress In Electromagnetics Research, PIER 123, 2012, pp.337-354.

[6] D. Sievenpiper, L. Zhang, R.F.J. Broas, N.G. Alexpolus, and E. Yablonovitch, "High-impedance electromagnetic surfaces with a forbidden frequency band, "IEEE Trans. Microwave Theory Tech., Vol. 47, 1999, pp. 2059-2074.

[7] M. Rezaei Abkenar, P. Rezaei, and R. Narimani, "Design of a slot-loaded EBG surface for a low profile antenna," IEEE AP-S Int. Symp. on Antennas and Propagation and USNC/URSI National Radio Science Meeting, July 2010.

[8] M. Rezaei Abkenar, and P. Rezaei, "Design of a compact slot-loaded EBG surface and its application in a low-profile antenna, "4th Int. Congress on Advanced Electromagnetic Materials in Microwaves and Optics, 2010, pp. 800-802.

[9] M. Rezaei Abkenar, and P. Rezaei, "Design of a novel EBG structure and its application for improving performance of a low profile antenna," 19th Iranian Conference on Electrical Engineering, 2011, pp. 3020-3024.

[10] M. Rezaei Abkenar, P. Rezaei, and R. Narimani, "Using mushroom-like EBG ground plane for improving radiation in low profile dipole antenna," 2010 URSI Int. Symp. on Electromagnetic Theory, 2010, pp. 635-638.

[11] "HFSS Release 11.0," Ansoft Corp., 2005.

[12] "CST Reference Manual." Darmstadt, Germany, Computer Simulation Technology, 2008.

[13] D.K. Cheng, Field and wave electromagnetics, 2nd Ed., Addison-Wesley, 1992.

[14] R. Collin, Field theory of guided waves, 2nd Ed., IEEE Press, New York, 1991.

[15] N. Ashcroft, and N. Mermin, Solid state physics, Saunders College Publishing, Orlando, 1976.

[16] C.A. Balanis, Antenna Theory: Analysis and Design, 3rd Ed., John Wiley & Sons, New York, 2005.

[17] S. Ramo, J.R. Whinnery, and T. Van Duzer, Fields and waves in communication electronics, 2nd Ed., John Wiley and Sons, New York, 1984.

[18] D. Sievenpiper, "High-impedance electromagnetic surfaces," Ph.D. dissertation at University of California, Los Angeles, 1999.
[19] F. Yang, and Y. Rahmat-Samii, "Polarization dependent electromagnetic band gap (PDEBG) structures: designs and applications," Microwave Optical Tech. Lett., Vol. 41, No. 6, 2004, pp. 439-444.
[20] S.M. Moghadasi, A.R. Attari, and M.M. Mirsalehi, "Compact and wideband 1-D mushroom-like EBG filters," Progress In Electromagnetics Research, PIER 83, 2008, pp. 323-333.
[21] L. Yang, Z.H. Feng, F.L. Chen, and M.Y. Fan, "A novel compact electromagnetic band-gap (EBG) structure and its application in microstrip antenna arrays," IEEE MTT-S Int. Microwave Symp. Dig., 2004, pp. 1635-1638.
[22] F. Xu, Z. X. Wang, X. Chen, and X.-A. Wang, "Dual Band-Notched UWB Antenna Based on Spiral Electromagnetic-Bandgap Strucure," Progress In Electromagnetics Research B, PIER B 39, 2012, pp. 393-409.
[23] Y. Gou, Q. Wu and J. Hua, "Design of A Bandwidth Extended and Compact Microstrip Antenna Based on CRLH TL with Slot Loaded," Microwave Conference Proceedings (APMC), 2011 Asia-Pacific, 2011, pp. 1762-1765.
[24] R. Remski, "Analysis of photonic band gap surfaces using Ansoft HFSS," Microwave Journal, Sept. 2000.
[25] F. Yang, and Y. Rahmat-Samii, "Reflection phase characterizations of the EBG ground plane for low profile wire antenna applications," IEEE Trans. Antennas Propagat., Vol. 51, No. 10, 2003, pp. 2691-2703.
[26] R. Gonzalo, P. Maagt, and M. Sorolla, "Enhanced Patch-Antenna Performance by Suppressing Surface Waves Using Photonic-Bandgap Substrates," IEEE Trans. Microwave Theory Tech., vol. 47, pp. 2131–8, 1999.
[27] F. Yang, K. Ma, Y. Qian, and T. Itoh, "A novel TEM waveguide using uniplanar compact photonic-bandgap (UC-PBG) structure," IEEE Trans. On Microwave Theory and Tech., Vol. 47, No. 11, 1999, pp. 2092-2098.
[28] W.L. Stutzman, and G.A. Thiele, Antenna theory and design, 2nd Ed., John Wiley and Sons, New York, 1998.

**Masoumeh Rezaei Abkenar** was born in Tehran, Iran, in 1983. She received the B.S. degree in Elecronic Engineering from Khaje Nasir Toosi University of Technology, Tehran, Iran, and the M.S. degree from Semnan University, Semnan, Iran, in 2006 and 2010, respectively. Her research interests include low-profile printed and patch antennas for wireless communication, miniaturized and multiband antennas, meta-materials and EBG structures interactions with antennas and RF passive components.

**Pejman Rezaei** was born in Tehran, Iran, in 1977. He received the B.S. degree in Electrical-Communication Engineering from Communication Faculty, Tehran, Iran, in 2000, and the M.S. and Ph.D. degrees from Tarbiat Modarres University, Tehran, Iran, in 2002 and 2007, respectively. Currently, he is assistant professor in the Semnan University, Semnan, Iran. His current research interests are Electromagnetics theory, metamaterial structure, theory and design of antenna, wave propagation, and satellite communication.

# A New Approach to Overcome the Count to Infinity Problem in DVR Protocol Based on HMM Modelling

Mehdi Golestanian*
Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
mehdi.golestanian@gmail.com
Reza Ghazizadeh
Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
r_ghazizadeh@yahoo.com

Received: 03/Mar/2013 Accepted: 11/Dec/2013

**Abstract**
Due to low complexity, power and bandwidth saving Distance Vector Routing has been introduced as one of the most popular dynamic routing protocol. However, this protocol has a serious drawback in practice called Count To Infinity problem or slow convergence. There are many proposed solutions in the literature to solve the problem, but all of these methods depend on the network topology, and impose much computational complexity to the network. In this paper, we introduce a new approach to solve the Count To Infinity using hidden markov model (HMM), which is one of the most important machine learning tools. As the modelling results show, the proposed method is completely independent from the network topology and simple with low computational complexity.

**Keywords:** Count To Infinity, Distance Vector Routing (DVR), Hidden Markov Model (HMM), Network Routing Protocol, Slow Convergence.

## 1. Introduction

At the beginning of the 1990s the wireless communication society witnessed a rapid growth of interest in mobile ad-hoc network (MANET) technology. Because of the dynamic behaviour of the MANETs, in order to improve the efficiency of communication networking, new strategies and protocols were required. MANETs utilize the conventional TCP/IP protocol [1,2]. However, the dynamic entity of these networks (i.e., mobility and resource constraints) needs the modification of TCP/IP protocol in different layers. One of the most challenging research topics of the MANETs is in the network layer known as MANETs routing. There are many routing protocols developed based on the properties of the MANETs. In packet-switched based communication theory, there are two major classes of routing protocols known as Link State Routing (LSR) and Distance Vector Routing (DVR) algorithms.

The link-state routing protocol, periodically broadcasts the link-state costs of its one hop neighbours to other nodes in a flooding manner. Each node in the network by receiving the update packets, updates its routing table. Using the link state information and employing a shortest path algorithm the LSR protocol determines the next hop node [3].

In the DVR protocol, each node maintains a routing table giving the shortest distance to each destination based on some metrics. Different metrics, such as number of hops, time delay, the number of packets queued in the routing queues, can be used to show the nodes distance. DVR protocol utilizes the Bellman-Ford algorithm to calculate the cost of each path, and chooses a path with the minimum cost as a best route to the destination. In DVR protocol, as each node makes its routing table according to information received from its neighbours, it provides less traffic, compared to LSR protocol which broadcasts the node information throughout the network in flooding manner. Furthermore, memory, bandwidth and power saving, are other prosperities that make the protocol more attractable routing protocol for network managers. Notwithstanding the mentioned benefits of the DVR, there is a serious drawback for this protocol called the *Count To Infinity* or slow convergence problem. The Bellman-Ford algorithm used in the DVR protocol does not prevent routing loops from happening and suffers from the count-to-infinity problem. When a node disconnects from the network due to a link failure, its neighbours cannot be informed about this *bad* news. Hence, they get the wrong routing information to reach the disconnected node. The wrong information propagates in the network slowly and all the node update their routing table for a long time until the routing table entity belongs to disconnected node reaches to an unacceptable amount (say infinite) according to the selected metrics in the protocol. In these conditions, the disconnected node information is removed. In section 3.2 this process is explained through an example.

There are several proposed techniques to overcome this drawback of the DVR protocol. However, all of the proposed methods in the literature are designed based on the topology of the network. This fact results in not completely solving of the problem for any random network topology. Furthermore, most of the proposed

* Corresponding Author

methods increase the complexity and computations of the routing algorithms (Section 2 is devoted to investigate some of these methods). In this paper, we introduce a new method to solve the Count to Infinity problem based on the concepts of system modelling and machine learning techniques. We use the Hidden Markov Model (HMM) as one of the most powerful machine learning tools in the field of system modelling to model the behaviour of the Count to Infinity problem in communication networks. After that we proposed a method to detect link failure in the network (which results in Count to Infinity) based on the structure of the HMM. The proposed method is completely independent of the network topology and the modelling results show that the proposed method can detect the Count to Infinity problem fast. Furthermore, using the HMM in routing managing of DVR protocol does not impose many computations to routers.

The rest of the paper is organized as follows. Section 2, reviews some related works about the Count to Infinity problem in DVR. In section 3 the Hidden Markov Model (HMM) and count to infinity problem are explained. The proposed method is introduced in section 4 and modelling and results are shown in the section 5. Section 6 presents some practical discussion about the proposed method and the last section concludes the paper.

## 2. Literature Review

There are several proposed methods to avoid count to infinity problem in literature [4,5,6,7]. However, all of them depend on network topology. [5] has proposed, CF-DVRP protocol to prepare backup set for each node when a link failure happens. A backup set is a set of nodes in the network selected to avoid updating the routing table of each node via loops in the network. However, in general, the availability of a backup set in a random network cannot be guaranteed. As mentioned in [6,8] four concepts, triggered update, split horizon, poison reverse and path hold-down have been suggested to solve count to infinity, but they did not solve this problem completely, due to their dependency to the network topology. To avoid routing loop, [6] classified possible self-passage or updating via loops, in nine different combinations of nodes. This classification and checking the self-passaging of each node can increase the complexity of algorithm and the delay of route selection in each router. In [7], slow convergence of networks is known the result of, Two-Node Loop Instability, Three-Node Loop Instability and more than Three-Node Instability, and proposed some solution to avoid the loops might integrate through groups of nodes. It is obvious that the same as [6], the main idea to prevent the Count to Infinity is based on classifying the network topology to some classes which might cause the updating the routing tables via loops. Based on their solution if any of these network topologies occurs in the network, the routing algorithm is aware that the Count to Infinity problem can happen in routing. Therefore, the routing scheme tries to avoid these topologies to find the path to destination. All these proposed methods are dependent on network topology, and cannot solve count to infinity problem when the topology changes. Also these methods cause message overhead and more complexity in routing tables by adding some parameters, adding sequence of numbers in header of each packet or using test packet to get information about topology of network [8,9]. As mentioned before, we use a new approach based on modelling the behaviour of the Count to Infinity problem. The main advantage of the proposed method is independency of the method to the network topology with a simple structure.

## 3. Background Information

In the two following subsections, some background information about the Hidden Markov Model (HMM) and the count to infinity problem in the DVR protocol is explained.

### 3.1 Hidden Markov Model (HMM)

To analyze a random sequence with the finite state structure, the Hidden Markov Model (HMM) is one of the most powerful machine learning techniques [10,11]. Generating random sequences which follows a certain pattern and detecting a particular pattern of random sequences are some of important applications of HMM model. The HMM can be considered as a generalization of a mixture model where the hidden variables which manage the mixture component to be selected for each observation, are related through Markov chains. Two main parts of the HMM are the hidden state and observation states. Observable sequence depends on the hidden states and each state has a probability distribution over the observation sequence which controls the interaction between these states. The only way to evaluate the sequence of hidden states is to observe the observation sequence. The transition between hidden states follows the Markov process in which the current state only depends on the previous state.

To construct a $(A,B,\pi)$ HMM model, first hidden and observation states, state transition probability matrix (A), in hidden states, observation sequence emission probability matrix (B) and initial vector ($\pi$) showing transitions start points are determined. Fig. 1 shows an example of HMM structure, with the initial vector, transition and emission probability matrices defined in the Equation 1. In this figure, two states S and T are the hidden states which interact via the transition probabilities (i.e. *P(S,T), P(T,S), P(S,S)* and *P(T,T)*). Furthermore, this figure shows the hidden states can emit different observations with different probabilities (i.e. *A, C* and *T*).

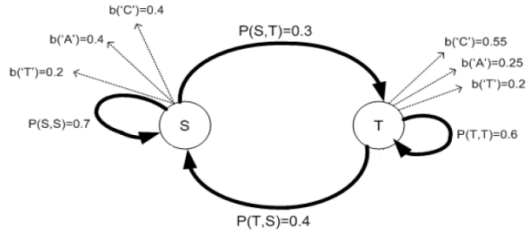



Fig. 1: An example of hidden states and observation sequences in HMM. S and T are hidden states, and C, A, T are observable option emitted from hidden states.

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}, B = \begin{pmatrix} 0.40 & 0.40 & 0.20 \\ 0.55 & 0.25 & 0.20 \end{pmatrix}, \pi = (1,0) \quad (1)$$

Estimation of transition and emission probability is one of the most important problems in HMM modelling. In fact the estimation of the HMM model expresses the behaviour and the hidden pattern in a random sequence. Therefore, in order to estimate the HMM parameters, HMM model is trained using enough numbers of observation sequences. Baum-Welch Algorithm is used to estimate transition and emission parameters. In this algorithm, if the number of observation sequences is too much (maybe infinite) acceptable results cannot be achieved and there is no guarantee that HMM can satisfy global results. In such a case, a heuristic algorithm is utilized to estimate the HMM parameters. Furthermore, transition and emission probabilities can be estimated based on the problem conditions without using any algorithm.

One of the most important usages of HMM model is in the decoding problems, which means HMM-decoding, computes the most probable sequence of the hidden states, according to the observation sequence, transition and emission probabilities. In other word, the HMM decoding determines if a random sequence can be emitted from the HMM model after enough training. In this paper, we use the HMM-decoding to introduce criteria to detect link failure in a network.

### 3.2 Count to Infinity Problem

In this subsection the Count to Infinity problem is explained through an example. Suppose in the network shown in Fig. 2, node A wants to join to the network. As it is illustrated in Fig. 2-a [6], after four information exchange, all nodes know their distance to the node A. Showing that the information of a new node propagates very fast throughout the network. However, when a node leaves the network or a link failure happens, it might take a long time that the nodes to be informed about a node disconnection. In the Fig. 2-b, suppose that the node A is disconnected. In the first step, node B updates its routing table entities according to the node Cs' routing table which has old information to reach the node A. Therefore, node B supposes that it can reach to the node A through the node C by the three hops. The node C and E distance information does not change. In second step according to routing table of the neighbours C the table entity in C is revised to four. A similar scenario continues, until the distance between A and other nodes reach to infinity, the definition of infinity depends on the metric and the number of the node in the network. Finally, nodes recognize the node disconnection and delete the node as information from their routing table [6].

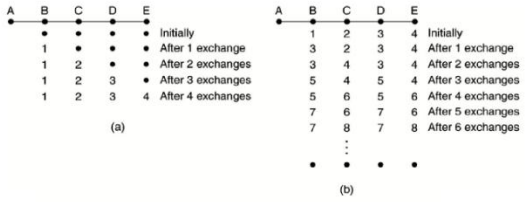



Fig. 2: (a) before and (b) after count to infinity problem.

It is obvious that the bad news, initiated due to a link failure, propagates very slowly in the networks. Hence, the nodes that are far from the broken link position are awarded very late resulting in weak network stability. The main reason of the slow convergence of the routing table in the DVR protocol is the table updating via loops that propagate the wrong routing information in the network. As mentioned there are several methods to avoid this wrong updating procedure which are so complex and not completely efficient. In the next section we present our approach to solve the problem based on the HMM-decoding with a simple structure and expandable to any network architecture.

## 4. Proposed Method

As mentioned in section 3.1 to establish the HMM model, first, two states of sequence, hidden and observation sequence, are defined. To model the Count to infinity problem, two hidden states are considered state 1, presents no disconnected node in the network and state 2 shows a node disconnection event moreover, the observation sequence is the sequence of hop numbers that is observed in the routing table showing the nodes distance which can change by topology changing. To simplify the observation sequence it is converted to a binary sequence, which identifies any changing in the amount of routing table entity. Therefore, after each updating, the amount of routing table entities is checked. If the certain entity has increased, the associated element in the binary observation sequence changes to 1, otherwise associated element is set to 0. In fact, by introducing binary observation sequence we try to model the behaviour of count to infinity by a string of ones and zeros. Therefore when series of ones are observed, the probability of link failure will increase.

In order to detect a link failure in the network the HMM-decoding is used to calculate the most probable hidden states sequence. It utilizes Viterbi algorithm according to observation sequence, transition and emission probabilities. Therefore HMM acts like a decision box and according to the binary observation sequence it finds any node disconnection by the HMM-

decoding. As the output of HMM-decoding is a probability of staying in each state, the decision making is a statistical model. In order to evaluate the method we run 100 different topologies with the introduced structure. The results are presented in the next section.

## 5. Modelling and Results

To show that how the proposed scheme can solve the count to infinity problem, suppose a network with 8 nodes which are connected to each other as shown in Fig. 3. Now, let us assume that the link between node A and B is broken and count to infinity problem happens.

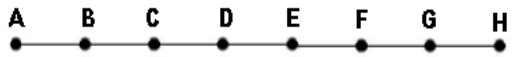
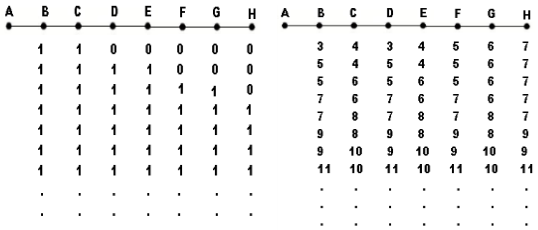



Fig. 3: Network with 8 nodes to simulate count to infinity problem

Fig. 4 denotes routing table and related binary observation sequence after breaking the link. Notice that the binary observation sequence shows the style of changing in the nodes distances. For example nodes B, D, F and H update their table at odd exchange times, as described in section 3.2.




(a) (b)

Fig. 4: (a) distance of each node to node A after each update and (b) related binary observation sequence.

To demonstrate a link failure in the network, the number of the binary observation sequence updates is limited by 3. This number in a large network should be larger. According to the updating times, 8 different possible observation sequences 000, 001, …, 111, determined by O1, O2, …, O8 respectively, are possible.

Fig. 5 illustrates the hidden states and observation sequences for the HMM model. To reduce the complexity of HMM model, observation sequence that leads to the same result in HMM-decode can be considered as one observation sequence, because these observation sequences present almost the same emission probability. Therefore, the observation sequences O2, O3, O5, O6 and O7, showing a normal changing in the network with no link failure, are assumed as one observation sequence. Fig. 6 shows the modified model in the assumed network.

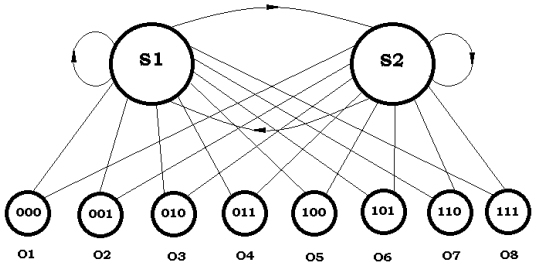



Fig. 5: Hidden states and observation sequences for HMM model.

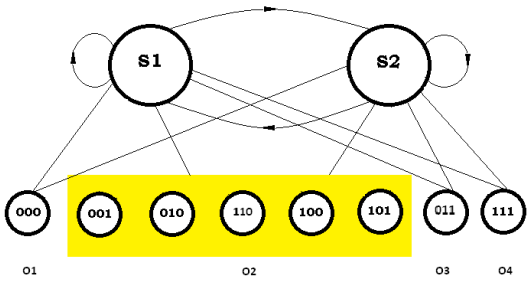



Fig. 6: Modified observation sequence for HMM model.

The next step is estimation of transition and emission probabilities. To estimate these parameters, a heuristic idea is employed. It is obvious that observation state O4 (in Fig. 6), showing 3 continuous increments in the updating table, is more likely to result count to infinity than the state 2, because, count to infinity is modelled with a string of many consecutive ones in observation sequence. Also O1, showing 3 continuous decrements or no changing in the routing table distances, has lower probability than the state 2 to result count to infinity. In this case, emission, transition and initial probability of HMM l are estimated as follows.

$$A = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 1 \end{pmatrix} \tag{2-1}$$

$$B = \begin{pmatrix} 0.5 & 0.20 & 0.15 & 0.15 \\ 0 & 0.15 & 0.30 & 0.55 \end{pmatrix} \tag{2-2}$$

$$\pi = \begin{pmatrix} 1 & 0 \end{pmatrix} \tag{2-3}$$

$\Theta=(A,B,\pi)$ shows HMM model of count to infinity problem. It is assumed that network starts at state 1.

As mentioned in the section 1, HMM provides a suitable criterion and a scale to evaluate the status of the network. Table I, shows the results of the HMM analysis which is evaluated according to the following procedure. After three updating, the status of the network is checked and the link failure probability is determined. To determine the current status of the network a threshold level is considered. If the output of HMM-decoding (probability of link failure) is less than 0.5, we assume that no link failure is occurred and normal changing is happened in the network topology. For probability of link failure more than 0.5 and less than 0.9 next three updating

and related binary observation sequence are checked to make sure that the increments is not due to the normal changing in the network topology when there is no link failure. Finally, for the probability more than 0.9, it can be assumed a link failure has occurred in the network. These threshold of decision making depends on HMM parameters introduced in Equations (2-1), (2-2) and (2-3).

Table 1: probability of link failure and status of the network based on binary observation sequence using HMM-decoding

| Time\node | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | … | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | … | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 3 | … | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Probability | … | 0.7857 | 0.7857 | 0.6667 | 0.6667 | 0.4286 | 0.4286 | 0 |
| Status | … | Wait and check next | Wait and check next | Wait and check next | Wait and check next | No link failure | No link failure | No link failure |
| 4 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Probability | … | 0.9683 | 0.9683 | 0.9483 | 0.9483 | 0.7857 | 0.7857 | 0.7857 |
| Status | … | Link failure | Link failure | Link failure | Link failure | Wait and check next | Wait and check next | Wait and check next |
| 7 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 9 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Probability | … | 0.9956 | 0.9956 | 0.9928 | 0.9928 | 0.9861 | 0.9861 | 0.9683 |
| Status | … | Link failure | Link failure | Link failure | Link failure | Link failure | Link failure | Link failure |

According to above discussion, the status of defined network in Fig. 3 is presented in the Table 1, however the slow convergence problem still exist. It means, it will be taken a long time for the nodes that are far from the link failure position to be aware of the bad news (in this example after 9 updating, the node H realizes that the node A is disconnected).

To solve this problem, the following equation is defined to compute a new probability (Pnew) based on the neighbours link failure probability.

$$P_{new} = \begin{cases} P_{old} & P_{old(ni)} < 0.5 \\ \alpha P_{old} + (1-\alpha) P_{old(ni)} & 0.5 < P_{old(ni)} < 0.9 \\ 0.5 + P_{old} & P_{old(ni)} > 0.9 \end{cases} \quad (3)$$

which P(ni) denotes the link failure probability of the neighbours and α is a constant number between 0 and 1 which is assumed α=(2/3). Each node checks the link failure probability of its neighbours. If there are two or more neighbours, maximum probability will be selected, and new probability is computed using Equation (3). For the decision making the same algorithm is performed as shown in Table 2. By using the modified probability defined in the Equation (3) if a link failure is detected the news propagates in the network very fast, which results in a fast convergence in the network. Table 3 shows the results based on the modified probability of the link failure.

Table 2: decision making based on Pnew

| Pnew | Status |
|---|---|
| 0 < Pnew < 0.5 | No link failure |
| 0.5 < Pnew < 0.9 | Wait and check next 3 update |
| 0.9 < Pnew < 1 | Link failure |

As it is obvious, according to Table 3, link failure between the node A and the network is detected very fast in comparison with results in the Table 1. Fig. 7 compares two defined models, model using HMM and model using modified probability, for the node H that is in the worst conditions to converge, because it has the biggest distance to the position of link failure. Fig. 8 denotes link failure probability for normal changing in the network topology (a random change is considered in the topology), where no link failure happens in network.

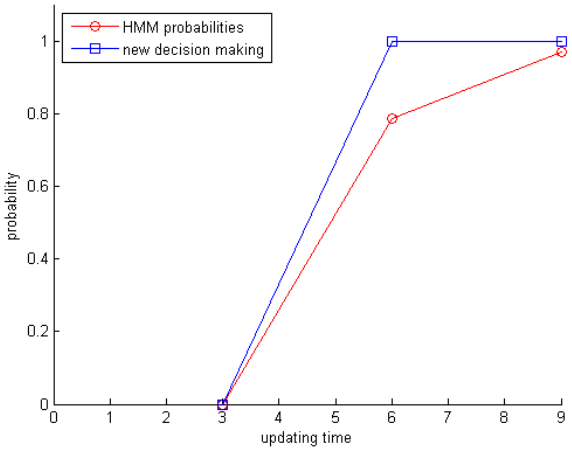



Fig. 7: Comparison of HMM probability and HMM using modified probability.

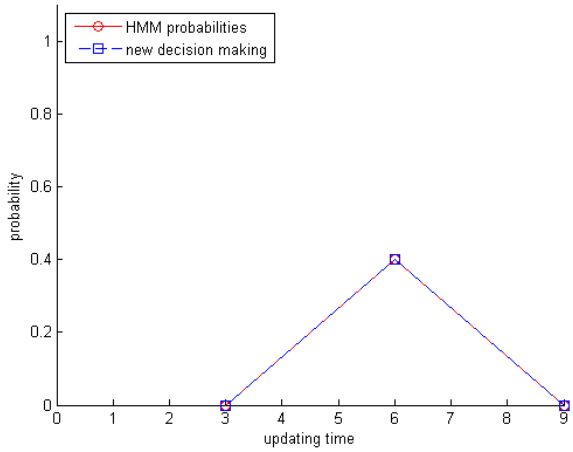



Fig. 8: HMM probability and HMM using modified probability when no link failure occurs.

Table 3: probability of link failure and status of network based on HMM-decoding and computation of modified probability in Equation (3)

| Time\node | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | … | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | … | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 3 | … | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Old_p | … | 0.7857 | 0.7857 | 0.6667 | 0.6667 | 0.4286 | 0.4286 | 0 |
| New_p | … | 0.7857 | 0.7857 | 0.7063 | 0.6666 | 0.5079 | 0.4286 | 0 |
| Status | … | Wait and check next | Wait and check next | Wait and check next | Wait and check next | No link failure | No link failure | No link failure |
| 4 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 5 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 6 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Old_p | … | 0.9683 | 0.9683 | 0.9483 | 0.7857 | 0.7857 | 0.7857 | 0.7857 |
| New_p | … | 0.9683 | 0.9683 | 1 | 1 | 1 | 1 | 1 |

| Time\node | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Status | … | link failure | Link failure | Link failure | Link failure | Link failure | Link failure | Link failure |
| 7 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 9 | … | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Old_p | … | 0.9683 | 0.9683 | 0.9683 | 0.9683 | 0.9683 | 0.9683 | 0.9683 |
| New_p | … | 0.9683 | 1 | 1 | 1 | 1 | 1 | 1 |
| Status | … | link failure | Link failure | Link failure | Link failure | Link failure | Link failure | Link failure |

According to Equation (3) and Table 2 the same results will be achieved for both methods when there is no link failure in the network. These results are presented in the defined network with 8 nodes. The results for larger network with more nodes can even be better because of using modified probability introduced in Equation (3). The probability of link failure propagates in network much faster than the news of link failure by other nodes.

## 6. Discussion

As we explained, our approach for fast convergence in the DVR protocol is to detect the specific pattern of routing information tables for any link failure in the network. At first step of the proposed method introduces a probability value for link failure based on the HMM concepts. After that the method by using the link failure probability of its neighbours can detect any link failure in the network so fast. One of the main advantages of the method is the independency from the network topology. Other proposed methods in the literature, to avoid the Count to Infinity problem perform some modifications on the network topology which could decrease the network performance. However, the proposed scheme detects the link failure in the network without any changing in the topology.

There are some practical points while using the proposed method which must be considered. One of the main issues of designing the HMM model is the interaction between the structure complexity and the link failure detection in the network. The number of the bits generating the observation sequences in the HMM model determines the structure complexity of the model and the accuracy of the model to detect the link failure. By increasing the number of the bits in the model the complexity of the model increases. For example, a 4-bit HMM model has 16 observation states. According to this number of states the training stage to determine the emission probabilities is more complex and cannot be computed easily. However, for larger numbers of bits generating the observation states, different status of the network can be detected with more accuracy. In other word, a 4-bit HMM model demonstrates that checking the link failure status of the network is performed with the accuracy twice the accuracy of a 2-bit HMM model. Therefore, in order to use the proposed model to detect the link failure effectively, one must be able to balance a trade-off between the complexity and the time accuracy of link failure detection. The following table shows different scenarios and how different number of bits can determine the time accuracy of the link failure in the network.

Table 4: the performance of the HMM model for different bit HMM modelling.

| Bit no. | State no. | Updating time to detect link failure | Numbers of node in the network |
|---|---|---|---|
| 2 | 4 | 8 | 8 |
| 3 | 8 | 6 | 8 |
| 4 | 16 | 2 | 8 |

As the modelling results demonstrate, 4-bit HMM model can detect the link failure faster than other model. The problem of high structure complexity can be solved by joining the observation states which have similar status for link failure in the network to one observation state. An example of joining is presented in Fig. 6 where states that have similar status for link failure are classified in one observation state. This joining can reduce the complexity of the model without decreasing the model performance.

## 7. Conclusion

In this paper, we introduced a new method to solve the slow convergence or Count To Infinity problem in DVR protocol. This model determines the existence of link failure based on a binary observation sequence. Count to infinity problem is modelled as certain behaviour of binary observation sequences based on nodes distances, therefore the other topology information like updating through loops or different form of node combination is not necessary. Since this method is independent of the network topology it can also detect multi-link failure easily. When multi-link failure happens in the network the status of link failure according to proposed model will be observed in routing table in number of broken link in the network. Therefore detecting link failure in such cases will be more easily and confident.

## References


[1] H. Bakht, "Survey of Routing Protocols for Mobile Ad-hoc Network," *International Journal of Information and Communication Technology Research*, vol.1, no.6, Oct.2011.

[2] J. Impson, J. Moody, N. Mehravari, "Network data routing in dynamic environments of potentially mobile nodes," in *Proc. SPIE communication and computing, vol.4741, no.70, pp.1-10,2002.*

[3] P. Manickam, T. Guru Baskar, M.Girija, D.Manimegalai, "Performance Comparisons of Routing Protocols in Mobile Ad-hoc Networks," *International Journal of Wireless & Mobile Networks*, vol.3, no.1, Feb.2011.

[4] K. Elmeleegy, A. L. Cox, and T. S. E. Ng, "On count to infinity induced forwarding loops Ethernet networks," in *Proc. 25th IEEE International Conference on Computer Communications, pp.1-13,2006.*

[5] B. Wang, Y. Guo, J. Zhou, L. Guan, "A fast convergence mechanism for distance-vector routing protocol," in *Proc ISECS International Colloquium on Computing, Communication, Control, and Management*. pp. 34-37, *2009*.

[6] A. J. Schmid, C. Steigner, "A solution for the counting to infinity problem of distance vector routing," *Technical report, Fachberichte Informatik, University Koblenz-Landau, D-56075 Koblenz, 1998. Available* http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.44.8729&rep=rep1&type=pdf

[7] A. D. Kothari, D. T. Patel, "Methodology to solve the count to infinity problem by accepting and forwarding correct and update information only using 'test' packet," in *Proc. IEEE International Advance Computing Conference (IACC 2009). pp.26-31,2009.*

[8] R.Kirutika , M.Phil, "An exploration of count to infinity problem in the networks," *International journal of Engineering and Science and Technology, vol.2, no.12, pp.7155-7159, 2010.*

[9] K. Elmeleegy, A. L. Cox, and T. S. E. Ng, "Understanding and Mitigating the Effects of Count to Infinity in Ethernet Networks," *IEEE Tran. Networking, vol.17, pp.186-199, 2009.*

[10] L. R. Rabiner, B. H. Juang, "An introduction to Hidden Markov Model," *IEEE ASSP Magazine, vol.3, pp.4-16, 1986.*

[11] L. R. Rabiner, "A toturial on hidden markov model and selected applications in speech recognition," in *Proc. IEEE, vol.77, no.2, pp.257-286, 1989.*



**Mehdi Golestanian** received the B.Sc. degree in electrical engineering from Shahid Bahonar University of Kerman, Iran, in 2010. And he received the master degree in communication systems from the Department of Electrical and Computer Engineering of Birjand University, Birjand, Iran, in 2012. His area research interests include Wireless communication networks, Digital Signal Processing, Bioinformatics, and Pattern recognition.

**Reza Ghazizadeh** received the B.Sc. and M.Sc. degrees in electrical engineering from Ferdowsi University of Mashhad, Iran, in 1992 and 1996, respectively, and the Ph.D. degree in Telecommunication engineering from the Southwest Jiatong university, China, in 2009.

In 1996, he joined the Department of electrical Engineering at University of Birjand, Iran as an instructor and was promoted to Assistant Professor in 2009. His research interests include the quality-of-service provisioning, radio resource management and, analysis and optimization of wireless communications networks.