<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>PAGE NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A New Method an Enhancement on Neural Cryptography with</td>
<td>1497</td>
</tr>
<tr>
<td>Multiple Transfers Functions and Learning Rules</td>
<td></td>
</tr>
<tr>
<td>N. Prabakaran, P. Loganathan and P. Vivekanandan</td>
<td></td>
</tr>
<tr>
<td>2. An Efficient Association Rule Mining for XML Data</td>
<td>1505</td>
</tr>
<tr>
<td>A. Bharathi, K. Anandakumar</td>
<td></td>
</tr>
<tr>
<td>3. Parallel Adaptive Temporal Prediction with Load Balancing</td>
<td>1514</td>
</tr>
<tr>
<td>For Fast Video Compression</td>
<td></td>
</tr>
<tr>
<td>S. Jeyakumar, S. Sundaravadivelu</td>
<td></td>
</tr>
<tr>
<td>4. Very Low Bit Rate Video Coding - A Review</td>
<td>1527</td>
</tr>
<tr>
<td>M. Thamarai, R. Shanmugalakshmi</td>
<td></td>
</tr>
<tr>
<td>5. Index Based Internet Traffic Sharing Analysis of</td>
<td>1539</td>
</tr>
<tr>
<td>Users By A Markov Chain Probability Model</td>
<td></td>
</tr>
<tr>
<td>D. Shukla, Sanjay Thakur</td>
<td></td>
</tr>
<tr>
<td>System</td>
<td></td>
</tr>
<tr>
<td>M. Lilly Florence, P. Balasubramanie</td>
<td></td>
</tr>
<tr>
<td>7. Generalized Neuron Based Digital Communication Channel Equalizer</td>
<td>1553</td>
</tr>
<tr>
<td>Vineeta Choudhary, B. K. Joshi</td>
<td></td>
</tr>
<tr>
<td>8. Traffic Analysis of Message Flow in Three Crossbar Architecture</td>
<td>1560</td>
</tr>
<tr>
<td>Space-Division Switches</td>
<td></td>
</tr>
<tr>
<td>D. Shukla, R. Singhai</td>
<td></td>
</tr>
<tr>
<td>9. Modelling and Dynamic Analysis of Pneumatic Control Valve with</td>
<td>1570</td>
</tr>
<tr>
<td>Stiction</td>
<td></td>
</tr>
<tr>
<td>S. Sivagamasundari and D. Sivakumar</td>
<td></td>
</tr>
<tr>
<td>10. Investigation of Image Fusion using Curvelet Transform</td>
<td>1579</td>
</tr>
<tr>
<td>Mary Praveena.S, Radhika.S and IlaVennila</td>
<td></td>
</tr>
</tbody>
</table>
A New Method an Enhancement on Neural Cryptography with Multiple Transfers Functions and Learning Rules

N. Prabakaran1                            P. Loganathan 2                       P. Vivekanandan3

ABSTRACT

The goal of any cryptographic system is the exchange of information among the intended users. We can generate a common secret key using neural networks and cryptography. Neural cryptography is based on a competition between attractive and repulsive stochastic forces. A feedback mechanism is added to neural cryptography which increases the repulsive forces. The partners A and B have to use a cryptographic key exchange protocol in order to generate a common secret key over the public channel. This can be achieved by two Tree Parity Machines (TPMs). In the proposed TPMs, each output vectors are compared, then updates from hidden unit using Hebbian Learning Rule, left-dynamic hidden unit using Random Walk Rule and right-dynamic hidden unit using Anti-Hebbian Rule with feedback mechanism. We can enhance the security of the system using different learning rules with different units. A network with feedback generates a pseudorandom bit sequence which can be used to encrypt and decrypt a secret message. In this paper, the most successful attack on neural cryptography is the majority flipping attack, which is presented here. The probability of a successful attack is calculated for different model parameters using numerical simulations.

Keywords : Neural Cryptography, Tree Parity Machine, Neural Synchronization, Feedback Synchronization, Majority Flipping Attacks.

1. INTRODUCTION

Neural cryptography is based on the effect that two neural networks are able to synchronize by mutual learning (Ruttor. A et al., 2006). In each step of this online learning procedure, they receive a common input pattern and calculate their output. Then, both neural networks use those outputs present by their partner to adjust their own weights. This process leads to fully synchronized weight vectors.

Synchronization of neural networks is, in fact, a complex dynamical process. The weights of the networks perform random walks, which are driven by a competition of attractive and repulsive stochastic forces. Two neural networks can increase the attractive effect of their moves by cooperating with each other. But, a third network is only listening to the communication. Therefore, bidirectional synchronization is much faster than unidirectional learning.

Two partners A and B want to exchange a secret message over a public channel. In order to protect the content against an attacker E, who is listening to the communication, A encrypts the message, but B needs A’s secret key for decryption. Without an additional private channel, A and B have to use a cryptographic
key exchange protocol in order to generate a common secret key over the public channel (Ruttor A et al., 2006). This can be achieved by synchronizing two TPMs, one for A and one for B, respectively. In this paper, we introduce a mechanism, which is based on the generation of inputs by feedback.

A measure of the security of the system is the probability $P_e$ that an attacking network is successful. We calculate $P_e$ obtained from the best known attack for different model parameters and search for scaling properties of the synchronization time as well as for the security measure. It turns out that feedback improves the security significantly but it also increases the effort to find the common key when this effort is kept constant, feedback only yields an improvement of security (Ruttor A et al., 2004).

In this paper, we analyze the influence of learning rules and order parameters on neural synchronization of TPMs are presented in Sec. 2. Also, we explain the synchronization of two TPMs with feedback mechanism and it is described in Sec. 3. Here, we show that the analysis of the security, probability of successful attack of the Majority Flipping Attacker with simulation results and it is explained in Sec. 4.

2. NEURAL SYNCHRONIZATION

The proposed Tree Parity Machine is used by partners and an attacker in neural cryptography consists of $K$-hidden units and $Y$-left-dynamic hidden units (Prabakaran N et al., 2008) and $Z$-right-dynamic hidden units, each of them being a perceptron with an $N$-dimensional weight vector $w_k$ (Godhavari T et al., 2005). When the hidden and dynamic hidden units receive an $N$-dimensional input vector $x_j$, these units produce the output bit.

The general structure of this network is shown in Fig.1. All inputs values are binary,

$$x_j \in \{-1,+1\}, \quad x_{im} \in \{-1,+1\}, \quad x_{ik} \in \{-1,+1\}$$

and the weights are discrete numbers between -L and +L,

$$w_q \in \{-L, -L+1, \ldots, -1, 1, \ldots, L\},$$
$$w_{im} \in \{-L, -L+1, \ldots, -1, 1, \ldots, L\},$$
$$w_{ik} \in \{-L, -L+1, \ldots, -1, 1, \ldots, L\}.$$  

where, $L$ is the depths of the weights of the networks.

In this paper, we analyze the influence of learning rules and order parameters on neural synchronization of TPMs are presented in Sec. 2. Also, we explain the synchronization of two TPMs with feedback mechanism and it is described in Sec. 3. Here, we show that the analysis of the security, probability of successful attack of the Majority Flipping Attacker with simulation results and it is explained in Sec. 4.

2. NEURAL SYNCHRONIZATION

The proposed Tree Parity Machine is used by partners and an attacker in neural cryptography consists of $K$-hidden units and $Y$-left-dynamic hidden units (Prabakaran N et al., 2008) and $Z$-right-dynamic hidden units, each of them being a perceptron with an $N$-dimensional weight vector $w_k$ (Godhavari T et al., 2005). When the hidden and dynamic hidden units receive an $N$-dimensional input vector $x_j$, these units produce the output bit.

Figure 1: A Structure of Tree Parity Machine with K=3, Y=3, Z=3 and N=4

$$\sigma_j = \text{sign} \left( \sum_{j=1}^{N} w_{ij} \cdot x_{ij} \right)$$  

$$\delta_i = \tanh \left( \sum_{m=1}^{N} w_{im} \cdot x_{im} \right)$$  

$$\gamma_i = \arctan \left( \sum_{k=1}^{N} w_{ik} \cdot x_{ik} \right)$$

where, equation (3) is the transfer function of the hidden unit (Kinzel W et al., 2000), the equation (4), is the transfer function of the left-dynamic hidden unit and
A New Method an Enhancement on Neural Cryptography with Multiple Transfers Functions and Learning Rules

(5), is the transfer function of the right-dynamic hidden unit.

The K-hidden units of $\sigma$, left-dynamic hidden units of $\delta$, and right-dynamic hidden units of $\gamma$ (Prabakaran N et al., 2008) define a common output bit $\tau$ of the total network and is given by

$$\beta_a = \prod_{i=1}^{K} \sigma_i$$

(6)

$$\beta_b = \prod_{i=1}^{K} \delta_i$$

(7)

$$\beta_c = \prod_{i=1}^{K} \gamma_i$$

(8)

where, equation (6) is output for the hidden units, equation (7) output for the left-dynamic hidden units and equation (8) output for the right-dynamic hidden units.

The two TPMs compare the output bits and then update the values between hidden units and dynamic hidden units as well as two parties A and B that are trying to synchronize their weights.

$$\psi_{i,A,B} = \text{comp}(\beta_a, \beta_b, \beta_c)$$

(9)

$$\phi_{i,A} = W_{i,j}^{A} x_{j}^{A} \tau^{A} \psi_{i}^{A}$$

(10)

$$\phi_{i,B} = W_{i,j}^{B} x_{j}^{B} \tau^{B} \psi_{i}^{B}$$

(11)

where, equation (9) represents comparison of the output of hidden, left-dynamic and right-dynamic hidden units. The equation (10) and (11) represent output of hidden, left and right-dynamic hidden units of A and B respectively.

Each of the two communication parties A and B has their own network with an identical TPM architecture. Each party selects a random initial weight vectors $w_i(A)$ and $w_i(B)$ at $t=0$.

2.1 Learning Rules

Both of the networks are trained by their mutual output bits $\tau^A$ and $\tau^B$. At each training step, the two networks receive common input vectors $x_i$ and the corresponding output bit $\tau$ of its partner (Ruttor A et al., 2004). We use the following learning rule:

(i) If the output bits are different, $\tau^A \neq \tau^B$, nothing is changed.

(ii) If $\tau^A = \tau^B = \tau$, the hidden and dynamic units are trained which have an output bit identical to the common output $\phi_{i,A,B}^A = \tau^{A/B}$.

(iii) To adjust the weights, we consider two different learning rules.

(a) Hebbian Learning rule for hidden units

$$w_i^A(t+1) = w_i^A(t) + x_i \tau^A \Theta(\tau^A \phi_i^A) \Theta(\tau^A \tau^B)$$

$$w_i^B(t+1) = w_i^B(t) + x_i \tau^B \Theta(\tau^A \phi_i^A) \Theta(\tau^B \tau^B)$$

(12)

where, $\Theta$ is the Heaviside step function (Engel A and Van Den Brock C, 2001), if the input is positive, then the output is 1 and if input is negative then the function evaluates to 0.

(b) Random walk learning for left-dynamic hidden units (Prabakaran N et al., 2008)

$$w_i^A(t+1) = w_i^A(t) + x_i \Theta(\tau^A \phi_i^A) \Theta(\tau^A \tau^B)$$

$$w_i^B(t+1) = w_i^B(t) + x_i \Theta(\tau^B \phi_i^B) \Theta(\tau^B \tau^B)$$

(13)

(c) Anti-Hebbian learning for right-dynamic hidden units

$$w_i^A(t+1) = w_i^A(t) - \phi_i x_i \Theta(\tau^A \phi_i^A) \Theta(\tau^B \tau^A)$$

$$w_i^B(t+1) = w_i^B(t) - \phi_i x_i \Theta(\tau^B \phi_i^B) \Theta(\tau^B \tau^B)$$

(14)
2.2 Order Parameters

The size of a matrix $F$ is $(2L+1) \times (2L+1)$ in TPMs. Their elements are $F(\mu)$, $F(\mu)$ and $F(\mu)$ where, ‘$\mu$’ is the state of the machines in the time step, ‘$i$’ is hidden units, where, ‘$j$’ is the left-dynamic hidden units and ‘$k$’ is the right-dynamic hidden units. The element $f_{ij}^{i}$ of matrix stands for the matching components in the $i$th weight-vector in which the A’s components are equal to ‘$q$’ and the matching components of B are equal to ‘$r$’. The element $f_{js}^{j}$, matching components of $j$th weight-vector in which the A’s components are equal to ‘$s$’ and the matching components of B are equal to ‘$t$’. The element $f_{uv}^{k}$, matching components of $k$th weight-vectors in which the A’s components are equal to ‘$u$’ and the matching components of B are equal to ‘$v$’. The values of $q$, $r$, $s$, $t$, $u$, $v$ are equal to -L, ..., -1, 0, 1, ..., L. The overlap of the weights belonging to the $i$th hidden unit, $j$th left-dynamic hidden unit and $k$th right-dynamic hidden unit in the two parties are given below:

\[
\begin{align*}
R_{i}^{i} &= \frac{w_{i}^{i} \cdot w_{i}^{i}}{N}, \quad R_{j}^{j} = \frac{w_{j}^{j} \cdot w_{j}^{j}}{N}, \quad R_{k}^{k} = \frac{w_{k}^{k} \cdot w_{k}^{k}}{N} \quad (15)
\end{align*}
\]

Also their norms $\rho_{i} = \frac{w_{i}^{i} \cdot w_{i}^{i}}{N}$, $\rho_{j} = \frac{w_{j}^{j} \cdot w_{j}^{j}}{N}$ and $\rho_{k} = \frac{w_{k}^{k} \cdot w_{k}^{k}}{N}$ are hidden, left and right-dynamic hidden units of A’s TPM respectively. $\rho_{i} = \frac{w_{i}^{i} \cdot w_{i}^{i}}{N}$, $\rho_{j} = \frac{w_{j}^{j} \cdot w_{j}^{j}}{N}$ and $\rho_{k} = \frac{w_{k}^{k} \cdot w_{k}^{k}}{N}$ are hidden, left and right-dynamic hidden units of B’s TPM respectively. They are given by the matrix elements

\[
R_{i}^{A,B} = \sum_{q,r} qr \cdot f_{qr}^{i} \quad (16)
\]

The equation (16-17) represent overlap between two hidden units, two left-dynamic hidden units and two right-dynamic hidden units of A and B respectively.

\[
R_{i}^{A,B} \quad (17)
\]

\[
R_{j}^{A,B} \quad (18)
\]

The equation (16-17) represent overlap between two hidden units, two left-dynamic hidden units and two right-dynamic hidden units of A and B respectively.

\[
R_{i}^{A,B} \quad (19)
\]

\[
R_{j}^{A,B} \quad (20)
\]

\[
R_{k}^{A,B} \quad (21)
\]

The equation (19-21) represent weight distribution of hidden units, left-dynamic hidden units and right-dynamic hidden units of A and B respectively.

These overlaps and norms fixed the probabilities of deriving the same internal representation via the normalized overlap,

\[
\rho = \frac{R_{ij}^{A,B}}{\sqrt{\rho_{i} \rho_{j}}} \quad \text{and} \quad \rho_{i}^{A,B} = \frac{R_{i}^{A,B}}{\sqrt{\rho_{i}^{A,B}}} \quad \text{then}
\]

\[
\rho_{i}^{A,B} = \frac{R_{i}^{A,B}}{\sqrt{\rho_{i} \rho_{j}}} + \frac{R_{j}^{A,B}}{\sqrt{\rho_{i} \rho_{j}}} + \frac{R_{k}^{A,B}}{\sqrt{\rho_{i} \rho_{j}}} \quad (22)
\]

More precisely, the probability of having different results in the $i$th hidden unit, $j$th left-dynamic hidden unit and $k$th right-dynamic hidden unit of the two parties is given by the well-known generalization error for the perceptron is given below

\[
\epsilon = \frac{1}{p} \cdot \arccos (\rho_{ijk}) \quad (23)
\]
The quantity $\epsilon_{\rho}$ is a measure of the distance between the weight vectors of the corresponding hidden units, left-dynamic hidden units and right-dynamic hidden units. Since the hidden unit, left-dynamic hidden unit, right-dynamic hidden units are independent, also the values $\epsilon_{\rho}$ determine the conditional probability $P_r$ for a repulsive step and $P_a$ for an attractive step between two hidden units, left-dynamic hidden units and right-dynamic hidden units given identical output bits of the two TPMs. In the case of identical distances $\epsilon_{\rho} = \epsilon$, the values of $K$, $Y$, and $Z$ are found as $K=3$, $Y=3$ and $Z=3$.

\[ P_r = \frac{1}{2} \frac{(1 - \epsilon)^2 + 3(1 - \epsilon)^3 \epsilon^2}{(1 - \epsilon)^2 + 9(1 - \epsilon)^3 \epsilon^2} \]  
\[ P_a = \frac{6(1 - \epsilon)^3 \epsilon^2}{3(1 - \epsilon)^2 + 9(1 - \epsilon)^3 \epsilon^2} \]  

The equation (24) and (25) represent probability of attractive and repulsive steps between two hidden units, two left-dynamic hidden units and two right-dynamic hidden units of A and B respectively.

3. SYNCHRONIZATION WITH FEEDBACK

The TPMs A and B start with different random weights and common random inputs. The feedback mechanism is defined as follows.

(i) After each step ‘t’ the input is shifted, $x_{i,j}(t + 1) = x_{i,j-1}(t)$ for $j > 1$.

(ii) If the output bits agree, $\tau^+(t) = \tau^-(t)$, the output of each hidden unit is used as a new input bit, $x_{i,j}(t + 1) = \phi(t)$ are set to common public random values.

(iii) After R steps with different outputs, $\tau^+(t) = \tau^-(t)$, all input vectors are reset to public common random vectors, $x_{i,j}^A(t + 1) = x_{i,j}^B(t + 1)$.

The feedback creates correlations between the weights and the inputs and therefore the system becomes sensitive to the learning rule. The components of the weights have a broad distribution in Anti-Hebbian Learning rule. The entropy per component is larger than 99% of the maximal value of $\ln(2L+1)$. For Hebbian or Random walk rule, the entropy is much smaller, because the values of the weights are pushed to the boundary values $\pm L$ (Ruttor A et al., 2004). Therefore, the network with the anti-Hebbian rule offers less information to an attack than the two other rules. In fig.2, we have numerically calculated the average synchronization time as a function of the number $L$ of components for the Hebbian rule of hidden units, Random walk rule of left-dynamic hidden units and Anti-Hebbian rule of right-dynamic hidden units. There is a large deviation from the scaling law $t_{\text{sync}} \propto L^2$ as observed for $R=0$.

Figure 2 : Average Synchronization Time Between Hebbian, Random Walk And Anti-Hebbian Of $T_{\text{sync}}$ And Its Standard Deviation As A Function Of $L$, From TPM With $K=3$. Simulation Results Obtained Using $N=10^3$
4. SIMULATION RESULTS

The attacker E tries to learn the weight vector of one of the two machines (Mislovaty R et al., 2004). The values of N and L are public as well as all the transmission through the channel: input $x_{ij}$ and output $\tau^{A/B}$. The information E lacks in each learning step are the values \(\{\sigma_i\}\) of A’s hidden units, \(\{\delta_i\}\) of A’s left-dynamic hidden units and \(\{\gamma_i\}\) of A’s right-dynamic hidden units. That is \(2^{K-1} \times 2^{Y-1} \times 2^{Z-1}\) possible updating scenarios A performs.

The most successful attack on neural cryptography is the Majority Flipping Attack, which is an extension of the Geometric Attack. The Attacker E uses an ensemble of ‘M’ TPMs. At the beginning, the weight vectors of all attacking networks are chosen randomly, so that the average initial overlap between them is zero. If the two partners A and B use queries for the neural key exchange, the success probability strongly depends on the parameter H. This can be used to regain security against the Majority Flipping Attack.

\[
P_{E} = \frac{1}{1 + e^{-\beta(H-\mu)}}
\]  

(26)

where, H is the absolute set value of the local field, \(\beta\) is sensitivity of \(P_{E}\) in regard of H and \(\mu\) is increases linearly with the synaptic depth. The two parameters \(\beta\) and \(\mu\) are a suitable fitting function for describing \(P_{E}\) as a function of H.

Here the probability of Majority Flipping Attack \(P_{flip(C)}\) decreases exponentially with increasing synaptic depth in Hebbian Learning Rule

\[
P_{flip(C)} \propto e^{-CL}
\]  

(27)

where, C increases linearly with R,

\[C = 0.09 + 1.06 \times 10^{-3} R\]

The probability of Majority Flipping Attack \(P_{flip(D)}\) in Random Walk Learning Rule

\[
P_{flip(D)} \propto e^{-DL}
\]  

(28)

where, D increases linearly with R,

\[D = 0.09088 + 1.06 \times 10^{-3} R\]

The probability of Majority Flipping Attack \(P_{flip(F)}\) in Anti-Hebbian Learning Rule

\[
P_{flip(F)} \propto e^{-FL}
\]  

(29)

where, F increases linearly with R

\[F = 0.09088 + 1.06 \times 10^{-3} R\]

The addition of three transfer functions are given by

\[
P_{flip(C+D+F)} \propto e^{-(C+D+F)L}
\]  

(30)

Figure 3: The Probability \(P_{flip(C+D+F)}\) As A Function Of L, Averaging 1000 Simulation With K=3, Y=3, Z =3 And M=1000 In Hebbian Learning Rule, Random Walk Learning Rule And Anti-Hebbian Learning Rule
We are able to predict the improvement of the security of success of attack from fig. 5. Also, we have show that the success probability of an attacker (P_e) is decreased using Hebbian learning rule for hidden unit, Random walk learning rule for dynamic unit and right-dynamic hidden unit for Anti-Hebbian learning rule in majority flipping attacks.

From the above results, we are able to identify that the feedback improves the security of neural cryptography. The synchronization time on the other side is also increased.

5. CONCLUSIONS
In the proposed TPMs, the synchronize time of the attacker is increased by the addition of three transfer functions for hidden unit using Hebbian learning rule, left-dynamic hidden unit using Random walk learning rule and right-dynamic hidden unit using Anti-Hebbian learning rule. A feedback mechanism for hidden, left-dynamic and right-dynamic hidden units has been increased the synchronization time of two networks and decreases the probability of a successful attack of the Majority Flipping Attack. The synchronization with feedback yields an improvement of the security of the system. The TPMs generate a secret key and also encrypt and decrypt a secret message.

REFERENCES


Author’s Biography

Vivekanandan Periyasamy received his Master of Science in Applied Mathematics from Madras University in 1978 and Doctor of Philosophy from Anna University in 1987. Also, he obtained his postgraduate degree in Master of Engineering in Computer Science and Engineering from Anna University in 1995. He is working as Professor of Mathematics, Department of Mathematics in Anna University from 1978. He visited Singapore, Malaysia, Japan, Bangladesh, Sultanate of Oman and USA for presenting research papers and Chairing Sessions. He has published more than 80 research papers in national and international journals. His areas of research are Neural Networks, Internet Security and Software Reliability.

Logathan.P, received M.Sc. degree from Presidency College, Chennai. He received his Ph.D. in Mathematics from Anna University. He is working as Assistant Professor in Department of Mathematics, Anna University. His areas of interest are Computational Fluid Dynamics, Heat and Mass transfer, Natural connection and Object Oriented Programming. He has published about ten papers in national and international conferences.

Prabakaran.N, received M.Sc in Computer Science from Anna University. He is doing research under the guidance of Dr. P. Vivekanandan. His areas of interest are network security, visual programming, computer networks and wireless. He had published five International Journals in national and international conferences.
An Efficient Association Rule Mining For XML Data

A.Bharathi\textsuperscript{1} K.AnandaKumar\textsuperscript{2}

\textbf{ABSTRACT}

XML association rule mining is an important problem in data mining domain. Currently, the problem of association rule mining on XML data has not been well studied. In this paper, we proposed an efficient association rule mining for large amount of XML data. The set of data is viewed as a binary table. The value of the itemset is one, if the corresponding XML data exist in the dataset, zero for otherwise. The frequent itemset is generated along with the candidate key. The closed itemset for the given data set is also generated. The closed itemset don’t have any superset. For both frequent itemset and closed itemset generation we use Apriori algorithm. The possible association rules are generated for the XML data. Then the generated Association rule is converted into XML format. Our proposed system EARM may reduce the memory storage size and it returns association rules with short response time.

\textbf{Keywords:} Association rules, Data Mining, Apriori Algorithm.

\textbf{1. INTRODUCTION}

Data Mining is the technique that is used to store and retrieve all types of data. During the mid- to late 1990s, commercial vendors began exploring the feasibility of applying traditional statistical and artificial intelligence analysis techniques to large databases for the purpose of discovering hidden data attributes, trends, and patterns. This exploration evolved into formal data-mining toolsets based on a wide collection of statistical analysis techniques. Data-mining techniques can generally be grouped into one of three categories: clustering, classifying, and predictive. The proliferation of database management systems has also contributed to recent massive gathering of all sorts of information. Today, we have far more information than we can handle: from business transactions and scientific data, to satellite pictures, text reports and military intelligence. Information retrieval is simply not enough anymore for decision-making. Confronted with huge collections of data, we have now created new needs to help us make better managerial choices. These needs are automatic summarization of data, extraction of the “essence” of information stored, and the discovery of patterns in raw data.

The goal of the project is to generate the association rules for XML data. This will reduce the memory storage size and it returns the association rules with short response time. The input XML document is given from which the frequent itemset and closed itemset is found using the apriori algorithm. The association rule is generated for frequent itemset and it is converted into XML format.
2. LITERATURE REVIEW

Data mining is a task of discovering interesting patterns from large amounts of data where the data can be stored in databases, data warehouses or other information repositories [1]. Data mining is the principle of sorting through large amounts of data and picking out relevant information. It is usually used by business intelligence organizations, and financial analysts, but it is increasingly used in the sciences to extract information from the enormous data sets generated by modern experimental and observational methods. It is a young field, drawing work from areas including database technology, artificial intelligence, neural networking, statistics, pattern recovery, knowledge acquisition and many other fields such as business, economics and bioinformatics.

2.1 Data Mining Categories

Clustering techniques group information based on a set of input patterns using an unsupervised or undirected algorithm. One example of clustering could be the analysis of business consumers for unknown attribute groupings.

Classifying techniques group or assign objects to predetermined groupings based on well-defined attributes. The groupings are often clusters discovered using the above techniques. An example would be assigning a consumer to a particular sales cluster based on their income level.

Predictive techniques take as input known attributes regarding a particular object or category and apply those attributes to another similar group to identify expected behavior or outcomes. For example, if a group of individuals wearing helmets and shoulder pads is known to be a football team, we can expect another group of individuals with helmets and pads to be a football team as well [2].

2.2 Data Mining Techniques

The following list describes many data-mining techniques in use today. Each of these techniques exists in several variations and can be applied to one or more of the categories above.

Regression Modeling: This technique applies standard statistics to data to prove or disprove a hypothesis. One example of this is linear regression, in which variables are measured against a standard or target variable path over time. A second example is logistic regression, where the probability of an event is predicted based on known values in correlation with the occurrence of prior similar events.

Visualization: This technique builds multidimensional graphs to allow a data analyst to decipher trends, patterns, or relationships.

Correlation: This technique identifies relationships between two or more variables in a data group.

Variance Analysis: This is a statistical technique to identify differences in mean values between a target or known variable and nondependent variables or variable groups.

Discriminate Analysis: This is a classification technique used to identify or “discriminate” the factors leading to membership within a grouping.

Forecasting: Forecasting techniques predict variable outcomes based on the known outcomes of past events.

Cluster Analysis: This technique reduces data instances to cluster groupings and then analyzes the attributes displayed by each group.
Decision Trees: Decision trees separate data based on sets of rules that can be described in “if-then-else” language.

Neural Networks: Neural networks are data models that are meant to simulate cognitive functions.

Apriori Algorithm

Data mining, more specifically the “Apriori Algorithm”, is used to derive association rules that represent relationships between input conditions and results of domain experiments. This enables the tool to answer questions such as “Given cooling medium and agitation during material heat treatment, predict cooling rate”. This allows users to perform case studies on the Web and use their results to optimize the involved processes, thus increasing customer satisfaction. Another interesting aspect is predicting material microstructure during heat treatment.

3. System Analysis

3.1 Existing System and Its Limitations

A first algorithm called Apriori was proposed, which generates (k+1) candidates using joins over frequent k-itemsets, must be generated by the algorithm. Although many of those frequent itemset may not be useful and may not exploit for finding association rules because some of these frequent itemset haven’t any interestingness antecedent or consequent in rules but generate them to find superior frequent itemset.

3.2 Proposed Systems

Input: Transactional data, minimum support and minimum confidence.

Output: Association rules between largest frequent itemsets. Change XML data to binary table form and count support of all frequent 1-itemsets. Remove the itemsets are not satisfied with user define minimum support. Repeatedly apply AND operation must find large frequent itemsets that can not be found. Logic XOR operation is applied to derive all interesting association rules between large frequent itemsets. Display association rules with xml format.

3.3 Advantage in the Proposed System

- Reduces the memory storage.
- Returns association rule with short response time.

4. Description

4.1 Module Analysis

Our proposed system consists of following modules:

1. Data Extraction Module
2. Data Conversion Module
3. Generation of Frequent Item set
4. Generation of Closed Item set
5. Association Rules Generation
6. Convert Rules to XML Module

4.1.1 Data Extraction Module

Sample XML document is given as an input in this module. Extraction of XML document using parser is obtained as an output to this module. The extracted data will be displayed in a notepad. This output is passed as an input to the second module.

4.1.2 Data Conversion Module

The extracted data from XML document is given as input to this module. The transactions were noted. The
extracted data is converted into binary format based on transaction items. If the item is present we will have a binary value of one and if the item is not present, the binary value is zero.

4.1.3 Generation of Frequent Item set

The binary formats of the transaction items are given as the input. The candidate and the frequent item sets are found for generating the Association Rules.

A frequent item set have the following characteristics:

* Any subset of a frequent item set must be also frequent
* A transaction containing {beer, diaper, nuts} also contains {beer, diaper}
* {beer, diaper, nuts} is frequent à {beer, diaper} must also be frequent
* No infrequent item set should be generated or tested.

Frequent Item set Property :

For generating the frequent item set, Apriori algorithm is used.

4.1.4 Generation of Closed Itemset

The binary formats of the transaction items are given as the input. The generation of closed itemset is the output to this module. The Mining frequent closed itemsets has the same power as mining the complete set of frequent itemsets. This reduces redundant rules to be generated and increases both efficiency and effectiveness of mining.

The idea behind this approach is to use conditional databases in a divide and conquer format. Given a list of all frequent items above minimum support, find conditional database of item in reverse order of support. Closed itemsets can be extracted iteratively from these conditional databases. Each conditional database may be further divided into more conditional databases.

4.1.5 Association Rules Generation

This module generates rules based on the support and confidence values from the frequent item sets. Association rule induction is a powerful method for so-called market basket analysis, which aims at finding regularities in the shopping behavior of customers of supermarkets, mail-order companies and the like. With the induction of association rules one tries to find sets of products that are frequently bought together, so that from the presence of certain products in a shopping cart one can infer (with a high probability) that certain other products are present. Such information, expressed in the form of rules, can often be used to increase the number of items sold, for instance, by appropriately arranging the products in the shelves of a supermarket (they may, for example, be placed adjacent to each other in order to invite even more customers to buy them together) or by
directly suggesting items to a customer, which may be of interest for him/her.

An association rule is a rule like “If a customer buys wine and bread, he often buys cheese, too.” It expresses an association between (sets of) items, which may be products of a supermarket or a mail-order company, special equipment options of a car, optional services offered by telecommunication companies etc. An association rule states that if we pick a customer at random and find out that he selected certain items (bought certain products, chose certain options etc.), we can be confident, quantified by a percentage, that he also selected certain other items. Association rule mining is always done by checking the regularities in data which can be done by asking questions like

1. What products were often purchased together? — Milk and Egg
2. What are the subsequent purchases after buying a PC?
3. What kinds of DNA are sensitive to this new drug?
4. Can we automatically classify web documents?
5. Some applications where Association rule mining is used broadly are
   5a. Basket data analysis, cross-marketing, catalog design, sale campaign analysis
   5b. Web log (click stream) analysis, DNA sequence analysis, etc.[6, 7]

The following is an example for generating Association rules. The table contains the Transaction ID and the items.[8]

<table>
<thead>
<tr>
<th>TID</th>
<th>Itemsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>T100</td>
<td>11,12,16</td>
</tr>
<tr>
<td>T200</td>
<td>12,14</td>
</tr>
<tr>
<td>T300</td>
<td>12,13</td>
</tr>
<tr>
<td>T400</td>
<td>11,12,14</td>
</tr>
<tr>
<td>T600</td>
<td>11,13</td>
</tr>
<tr>
<td>T600</td>
<td>12,13</td>
</tr>
<tr>
<td>T700</td>
<td>11,13</td>
</tr>
<tr>
<td>T800</td>
<td>11,12,13,16</td>
</tr>
<tr>
<td>T900</td>
<td>11,12,13</td>
</tr>
</tbody>
</table>

Let us consider the 3-Itemset {11, 12, 16} with support of 2%. The generation of association rules for this itemset is as follows

I1 $\rightarrow$ I2 $\Rightarrow$ I6 $\Rightarrow$ confidence=2/4=60%
I1 $\rightarrow$ I6 $\Rightarrow$ I2 $\Rightarrow$ confidence=2/2=100%
I2 $\rightarrow$ I6 $\Rightarrow$ I1 $\Rightarrow$ confidence=2/2=100%
I1 $\rightarrow$ I2 $\Rightarrow$ I6 $\Rightarrow$ confidence=2/6=33%
I2 $\rightarrow$ I1 $\Rightarrow$ I6 $\Rightarrow$ confidence=2/7=29%
I6 $\rightarrow$ I1 $\Rightarrow$ I2 $\Rightarrow$ confidence=2/2=100%

4.1.6 Convert Rules to XM

An XML file is called a document which has one top-level Element. The elements much match the start and end tags or a combined tag. The content between the tags can be a text or an element. The attributes are the Name/Value pairs with the Start or combined tag.[4]

XML and HTML Difference

In HTML, both the tag semantics and the tag set are fixed. An <h1> is always a first level heading and the tag <ati.product.code> is meaningless. The W3C, in conjunction with browser vendors and the WWW community, is constantly working to extend the definition of HTML to allow new tags to keep pace with changing technology and to bring variations in presentation (stylesheets) to the Web. However, these changes are
always rigidly confined by what the browser vendors have implemented and by the fact that backward compatibility is paramount. And for people who want to disseminate information widely, features supported by only the latest releases of Netscape and Internet Explorer are not useful.

XML specifies neither semantics nor a tag set. In fact XML is really a meta-language for describing markup languages. In other words, XML provides a facility to define tags and the structural relationships between them. Since there’s no predefined tag set, there can’t be any preconceived semantics. All of the semantics of an XML document will either be defined by the applications that process them or by stylesheets.[5]

**XML and SGML Differences**

XML is defined as an application profile of SGML. SGML is the Standard Generalized Markup Language defined by ISO 8879. SGML has been the standard, vendor-independent way to maintain repositories of structured documentation for more than a decade, but it is not well suited to serving documents over the web (for a number of technical reasons beyond the scope of this article). Defining XML as an application profile of SGML means that any fully conformant SGML system will be able to read XML documents. However, using and understanding XML documents does not require a system that is capable of understanding the full generality of SGML. XML is, roughly speaking, a restricted form of SGML.[6]

For technical purists, it’s important to note that there may also be subtle differences between documents as understood by XML systems and those same documents as understood by SGML systems. In particular, treatment of white space immediately adjacent to tags may be different.

**Why XML?**

In order to appreciate XML, it is important to understand why it was created. XML was created so that richly structured documents could be used over the web. The only viable alternatives, HTML and SGML, are not practical for this purpose.

HTML, as we’ve already discussed, comes bound with a set of semantics and does not provide arbitrary structure. SGML provides arbitrary structure, but is too difficult to implement just for a web browser. Full SGML systems solve large, complex problems that justify their expense. Viewing structured documents sent over the web rarely carries such justification.

This is not to say that XML can be expected to completely replace SGML. While XML is being designed to deliver structured content over the web, some of the very features it lacks to make this practical, make SGML a more satisfactory solution for the creation and long-time storage of complex documents. In many organizations, filtering SGML to XML will be the standard procedure for web delivery. [3]

**5. Screen Shots**

**5.1 Data Extraction from XML Document**

Figure 5.1: Data Extraction from XML Document
5.2 Data Conversion

5.3 Generate Frequent Itemset

5.4 Generate Closed Itemset

5.5 Association Rule Generation
5.6 Convert Rules To XML Format

Figure 5.6: Convert Rules to XML Format

6. Conclusion

The proposed system for mining association rules for XML data has been successfully implemented. This proposed algorithm may reduce storage memory size. The user-friendly interface designed for this paper can enable anyone to learn about the association rule mining for XML databases.

7. Future Enhancement

Currently, this system uses static XML data and the algorithm uses text file as input. This can be extended to generate output for the dynamic XML data and from the XML databases. The system can be tested with the real-time data in the future for improving it’s accuracy in generating association rules. Since, this is an ongoing research work in the area of data mining; any open issues can be addressed in the future.

REFERENCES


**Author’s Biography**

A. Bharathi received her Bachelor of Engineering Degree from Kongu Engineering College in 1998, Perundurai, Master of Engineering Degree from Bannari Amman Institute of Technology, Sathyamangalam, in 2007 and she is doing Doctor of Philosophy in Computer Science and Engineering from Anna University, Coimbatore. She had 12 years of teaching experience. Currently she is working as Assistant Professor, Department of IT, Bannari Amman Institute of Technology, and Sathyamangalam. Her professional activities include Guided Ten UG projects and guiding Seven UG and Three PG projects. She has presented 9 papers in International and National Conferences. She is published 3 national and 1 international journals.

K. Ananda Kumar was born in Tamil Nadu, India on March 1975. He received the B.Sc Degree in Physics from Bharathiar University in 1995. He received his MCA Degree in Computer Applications from Bharathiar University in 1998. He received his M.Phil from Periyar University in 2006 and he is doing Doctor of Philosophy in Computer Science and Engineering from Bharathiar University, Coimbatore. He had 12 years of teaching experience. Currently he is working as HOD in Computer Applications Department, Dr. SNS Rajalakshmi College of Arts and Science College, Coimbatore. His professional activities include Guided Fifteen PG projects and Five M.Phil and guiding Seven PG and Three M.Phil projects. Published and presented 4 papers in International and National journals and also Conferences.
Parallel Adaptive Temporal Prediction with Load Balancing for Fast Video Compression

S. Jeyakumar¹  
S. Sundaravadivelu ²

ABSTRACT

Video image compression has been an area where the computational demand is far above the capacity of conventional sequential processing. In this paper, we present a parallel adaptive motion estimation model for video compression using cluster computing on a local network with balanced load. The method used for temporal prediction is adaptive, in the manner in which, frames with very few motion changes are predicted in its integer wavelet domain and for high motion activity frame, motion compensation is applied in its spatial domain. This approach gives good compression rate. Secondly we apply a parallel compression model by having a multiple networked heterogeneous personal computer systems that perform compression on different input frames simultaneously. Also computing load is distributed properly among all processors by resource management technique of cluster computing. The implementation result shows that the proposed parallel method has better speedup than sequential algorithm and very much suitable for online video applications.

Keywords: Data Parallelism, Task Parallelism, Motion Vector, Temporal Predictor, Message Passing Interface.

1. INTRODUCTION

Video compression is a common need in today’s multimedia and Internet world. The increased used of video data in telecommunications services, multimedia applications, the corporate environment, the entertainment industry, and at home has made digital video technology a necessity. Video images are moving pictures which are sampled at frequent intervals usually, 25 frames per second and stored as sequence of frames. A problem, however, is that digital video data rates are very large, typically in the range of 150 Mbits/sec. Data rates of this magnitude would consume a lot of the bandwidth, storage and computing resources in the typical personal computer. For this reason, video compression standards have been developed and intensive research is going on to develop effective techniques to eliminate picture redundancy, allowing video information to be transmitted and stored in a compact and efficient manner [6]. A video image consists of a time-ordered sequence of frames of still images. In video streams, adjacent frames tend to be very similar. MPEG compression makes use of this temporal redundancy of the data and allows video to be compressed using motion estimation. There are 2 kinds of frames defined in the video stream, each of which is compressed differently: Intra-frames (I-frames) and Inter-frames (P-frames). I-frames are treated as independent images, with no reference to any past frames. The encoding scheme used for I-frames is similar to JPEG LS compression,
where as P-frames are treated as predicted frames. An obvious solution to such frames would be predictive coding based on previous or future frames and the compression proceeds by coding the residual error [3],[8].

Though video compression system is available and several advanced compression techniques evolved, the compression tasks are computationally intensive, mainly due to the large amount of data to be processed and the time consuming repetitive operation of the processing algorithms. One of the ironies to come out of image compression research is that as the data rates come down where as the computational complexity of the algorithms increases. This leads to the problem of long execution times to compress an image sequence. It is apparent that, in order to transmit or store real time video, some scheme for fast video compression using parallel processing is necessary [9][10]. One of the significant characteristics of video compression algorithms, that make them very attractive for the use of parallel processing techniques, is that, the motion estimation algorithms can operate on different frames simultaneously, with each of these frames being coded separately. Hence frame coding facilitates making the image compression algorithm, adaptive to local image statistics and then be performed in parallel [13], [15].

The rest of the paper is organized as follows: A survey on the related work is presented in Section 2. Section 3, describes the proposed adaptive temporal prediction method. In Section 4, we present the parallel implementation model of proposed technique. Detailed experimental results and discussion have been given in Section 5, and finally, conclusions are drawn in Section 6.

2. RELATED WORKS

Video coding with good compression rate is useful for many real time applications, such as telemedicine, video conferencing and other multimedia systems [8].

Brunello et al. [2], introduced a temporal prediction technique based on block motion compensation and an optimal 3-dimensional linear prediction algorithm. In their scheme, based on motion information, the pixel to be coded is predicted by a linear combination of neighboring pixels in the current and reference frames. However, this method has more computation than reconstruction.


A wavelet-based lossless video coding algorithm is proposed by Gong et al. [7]. In this approach, block motion compensation was first performed in the spatial domain, and then wavelet coefficients of the prediction residuals were coded and transmitted.

In [14], Ying et al., proposed an enhanced adaptive pixel based predictor which exploits the motion information among adjacent frames using extremely low side information. However, since the prediction is pixel based, the computational complexity is very high which may not be suitable for online real time applications.

In this paper, in order to improve the compression efficiency, we propose that, the motion vectors are adaptively estimated in spatial and wavelet domain based on inter frame similarity using correlation approach. For improving the speed up of motion estimation, parallel processing is to be applied. Generally, approaches used for parallelism can largely be divided into two major
areas: architecture-driven approach and algorithm-driven approach [4],[10]. The first is the use of special purpose architectures designed specifically for performing operations in parallel, such as an array of DSP chips to implement JPEG and MPEG, high performance parallel computers and high speed networking. The second approach is algorithm driven, in which the structure of the compression algorithm is implemented with data and task parallelism [4]. In this paper we implement a parallel approach by distributing the work on a network of PCs.

3. Proposed Temporal Prediction

The objective of this work is to study the relationship between the operational domains for prediction, according to temporal redundancies between the sequences to be encoded. Based on the motion characteristics of the inter frames, the system will adaptively select the spatial domain or wavelet domain for prediction. The block diagram of proposed method is shown in Figure 1.

3.1 Adaptive Domain Selection

This step aims to determine the operational mode of video sequence compression according to its motion characteristics. The candidate operational modes are spatial domain and wavelet domain. The wavelet domain is extensively used for compression due to its excellent energy compaction. However, Gong et al [7] pointed out that motion estimation in the wavelet domain might be inefficient due to shift invariant properties of wavelet transform. Hence, it is wise to predict all kinds of video sequences in the spatial domain alone or in the wavelet domain alone. Hence a method is introduced to determine the prediction mode of a video sequence adaptively according to its temporal redundancies. The amount of temporal redundancy is estimated by the inter frame correlation coefficients of the test video sequence [18].

The inter frame correlation coefficient between frames can be calculated by (1). If the inter frame correlation coefficients are smaller than a predefined threshold, then the sequence is likely to be a high motion video sequence.

In this case, motion compensation and coding the temporal prediction residuals in wavelet domain would be inefficient; therefore, it is wise to operate on the sequence in the spatial mode. Those sequences that have larger inter frame correlation coefficients are predicted in direct spatial domain. The frames that have more similarities with very few motion changes are coded using temporal prediction in integer wavelet domain.

\[ C_{\text{var}} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ij} - \bar{x}) (y_{ij} - \bar{y})}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ij} - \bar{x})^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (y_{ij} - \bar{y})^2\right)^{\frac{1}{2}}} \]  

Figure 1: Block Diagram Of Proposed Method
3.2 Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) is the most popular transform for image-based application [14]. A 2-dimensional wavelet transform is applied to the original image in order to decompose it into a series of filtered sub band images. At the top left of the image is a low-pass filtered version of the original and moving to the bottom right, each component contains progressively higher-frequency information that adds the detail of the image. It is clear that the higher-frequency components are relatively sparse, i.e., many of the coefficients in these components are zero or insignificant. The wavelet transform is thus an efficient way of decorrelating or concentrating the important information into a few significant coefficients. The wavelet transform is particularly effective for still image compression and has been adopted as part of the JPEG 2000 standard and for still image texture coding in the MPEG-4 standard. Figure 2 shows the representation of DWT sub bands of a three level multi resolution decomposition.

\[
\begin{array}{cccc}
LL_1 & HL_3 & & \\
& LH_3 & HH_3 & \\
LH_2 & & HH_2 & \\
& LH_1 & & HH_1 \\
\end{array}
\]

Figure 2 : DWT Sub Bands

The Haar wavelet is the first known wavelet and was proposed in 1909 by Alfréd Haar. The Haar wavelet is the simplest possible wavelet with coefficients [0.707, 0.707]. The S transform is the integer version of the Haar transform [14] which has the lowest computational complexity, and reasonably well both for lossy and lossless compression. The forward S transform equations are given in (2).

\[
h(i) = x(2i + 1) - x(2i) \\
r(i) = x(2i) + \frac{h(i)}{2}
\]

where \(x(i)\) is the input signal, \(h(i)\) is the high frequency sub-band signal and \(r(i)\) is the low-frequency sub-band signal.

3.3 Temporal Residual Prediction

Motion estimation obtains the motion information by finding the motion field between the reference frame and the current frame. It exploits temporal redundancy of video sequence, and, as a result, the required storage or transmission bandwidth is reduced by a factor of four. Block matching is one of the most popular and time-consuming methods of motion estimation [2],[3]. This method compares blocks of each frame with the blocks of its next frame to compute a motion vector for each block; therefore, the next frame can be generated using the current frame and the motion vectors for each block of the frame.

Block matching algorithm is one of the simplest motion estimation techniques that compare one block of the current frame with all of the blocks of the next frame to decide where the matching block is located [8]. Considering the number of computations that has to be done for each motion vector, each frame of the video is partitioned into search windows of size \(H\times W\) pixels. Each search window is then divided into smaller macro blocks of size \(8\times 8\) or \(16\times 16\) pixels. To calculate the motion vectors, each block of the current frame must be compared to all of the blocks of the next frame with in the search range and the Mean Absolute Difference
(MAD) for each matching block is calculated using equation (1)

$$MAD_{n}(x, y)=\sum_{i=0}^{N} \sum_{j=0}^{N}[x(i, j)-y(i+m, j+n)]$$

(3)

where N*N is the block size, $x(i,j)$ is the pixel values of current frame at $(i,j)$ th position and $y(i+m,j+n)$ is the pixel value of reference frame at $(i+m,j+n)$ th position. The target mapping of current frame and reference with in the search range is shown in figure 3.

![Figure 3: Inter Frame Target Window Mapping](image)

The block with the minimum value of the Mean Absolute Difference (MAD) is the preferred matching block. The location of that block is the motion displacement vector for that block in current frame. The best motion vector for the target window with the minimum MAD is determined by,

$$(m_0, n_0) = \{\text{minimum MAD (Tw)}\}$$

(4)

where $(m_0, n_0)$ indicates the motion displacement of the target window Tw. Then the temporal predictor of pixel $p(x,y)$ can be obtained by,

$$\hat{p}^*(x, y)=p(i+m_0, j+n_0)$$

(5)

and the temporal prediction residual is

$$e_2 = p(x, y) - \hat{p}^*(x, y)$$

(6)

### 3.4 Coding the Prediction Residual

The temporal prediction residuals from adaptive prediction are encoded using Huffman codes. Huffman codes are used for data compression that will use a variable length code instead of a fixed length code, with fewer bits to store the common characters, and more bits to store the rare characters. The idea is that the frequently occurring symbols are assigned short codes and symbols with less frequency are coded using more bits. The Huffman code can be constructed using a tree. The probability of each intensity level is computed and a column of intensity level with descending probabilities is created. The intensities of this column constitute the levels of Huffman code tree. At each step the two tree nodes having minimal probabilities are connected to form an intermediate node. The probability assigned to this node is the sum of probabilities of the two branches. The procedure is repeated until all branches are used and the probability sum is 1. Each edge in the binary tree, represents either 0 or 1, and each leaf corresponds to the sequence of 0s and 1s traversed to reach a particular code. Since no prefix is shared, all legal codes are at the leaves, and decoding a string means following edges, according to the sequence of 0s and 1s in the string, until a leaf is reached.

The code words are constructed by traversing the tree from root to its leaves. At each level 0 is assigned to the top branch and 1 to the bottom branch. This procedure is repeated until all the tree leaves are reached. Each leaf corresponds to a unique intensity level. The codeword for each intensity level consists of 0s and 1s that exist in the path from the root to the specific leaf.

### 4. Implementation Of Parallel Model

In this paper, we introduce parallel image processing approach that applies distributed client-server computing concept. This technique uses the power of local computer network with master-slave concept [5],[9],[12]. The environment consists of a server with a number of workstations. A simple master-slave computing model is shown in the figure 4.
This is a simple approach which distributes independent, non-overlapping image blocks on a multi, single processor cluster of workstations, using Message Passing Interface mechanism. The master-slave concept is the standard approach, in which the master sends the data to a slave and the slave sends back the output after computation [13]. The system consists of a master server and many slave processors. The master server controls synchronization of all processes, assigns slave processes which blocks to process, notifies the output server (same machine or another client) which blocks are done. The output server combines output blocks to a output file and slave processes perform processing and outputs results, notifies master process which and when the assigned blocks are done.

4.1 Parallel Temporal Prediction

The block diagram of the proposed parallel implementation of motion vector estimation and motion vector approximation is shown in figure 6.

The algorithm and steps for our parallel work is as below.

a. Maintain a collaborative memory in the master server to keep the video sequences as a queue [5].
b. By default, the first and last frames are considered as key frames and compressed using spatial predictor method before the parallel run.

c. Server then distributes the frames among all the processors in the cluster of workstations sequentially.
d. Each slave then proceeds through its assigned frame, performs the temporal prediction adaptively.
e. After finishing up, the slave sends message to master server and it writes 1 to the completion register of the respective frames.
f. When the completion register of specific workstation becomes 1, the processed frame is sent to the output server from the slave processor.
The master server computes the next chunk of frames for this slave to compress, encode a message and send it to the slave.

The process is repeated and the server will remain in the waiting state till the completion of all slave processes.

### 4.2 Load Balancing

The key to good parallelization is, how to synchronize the slave processes, so that no slave process is kept waiting for assignment from the master to get frames to compress and that output server is not waiting for a slower slave process, for the output frames to generate the compressed file.

Load balancing is the major criterion to improve throughput or speed up execution of the set of jobs while maintaining high processor utilization. Load balancing is the allocation of the workload among a set of co-operating nodes [1],[6],[15]. Generally, when working with heterogeneous nodes in a cluster, the computing powers of the intervening nodes are a factor used to analyze the distribution of the work to be done. In task parallelism, if the type of work is static, a predictive load balancing function can be used. It is defined as:

\[
D = F (P_i, T_i) \tag{7}
\]

where \(D\) is the distribution function, \(P_i\) is the computing power of processor \(i\) and \(T_i\) the total work. The total workload \(T_i\) will be allocated to the \(N\) processor at the moment of starting the application, according to the distribution function \(D\). In parallel motion estimation for video compression, a group of frames is assigned to a particular node depending on the computing power available for that node. The typical computing resources consist of CPU Clock speed in MHz and memory capacity in Megabyte (MB) or Gigabyte (GB). This speed and memory characteristics then can be used to distribute the computing load evenly. Hence, for data distribution, the number of frames assigned to a node is related to the computing power of that node relative with the total computing power of nodes.

This can be expressed as:

\[
N_f \cdot \frac{(W_i + W_m)}{(W_i + W_m)} - T_f \sum \frac{W_i}{W_i + W_m}, W_m \tag{8}
\]

where \(W_i, W_m\) and \(T_f\) are the number of frames assigned to node \(i\), weight of node \(i\) according to its speed and memory capacity and total number of frames respectively. With weighted round robin, the loads are distributed among the participating nodes on a round robin fashion. The capacity information of each node is collected and sent to master node by the resource-monitoring system before start of assigning frames.

During each frame processing the master node distributes the frame accordingly, and this process will continue until all frames are processed.

### 5. Results and Discussion

The proposed compression algorithm has been implemented in Java platform for standard test video sequences. In order to check the efficiency of our proposed adaptive temporal prediction and its parallel implementation, 50 video frames (Tennis sequences) were processed with 4 slaves and one master server. By default, first and last frames are considered as I-frames and remaining 48 frames are compressed using block based temporal prediction. The parallel output is verified with sequential processing and in all cases the output matches with the output produced by sequential version.

The experiments were done on a cluster architecture consisting of 1 server 4 nodes:
- Server - HP Proliant Server MI 350
- Slave #1 - Pentium IV 2.66 GHz 512 MB RAM
Parallel Adaptive Temporal Prediction with Load Balancing for Fast Video Compression

- Slave #2 - Pentium Quad Core 2.4 GHz 1 GB RAM
- Slave #3 - Pentium D CPU 2.8 GHz 480 MB RAM.
- Slave #4 - Pentium IV 2.4 GHz of 256 MB RAM

The systems are interconnected to one 24-port 10/100 mbps switch using CAT 5E UTP cable. The prediction results for the Tennis sequence frames 1 and 3 are shown in figures 7 and 8 respectively. Frame 1 has high motion characteristics from frame 0 and hence its motion vectors are predicted and residuals are compressed in spatial domain. Frame 3 more similarity with frame 2 and its motion estimation is done in wavelet domain.

Figure 7: Prediction for Tennis Frame 1 in Spatial Domain
(a. Frame 0  b. Reconstructed Frame 1)

Figure 8: Prediction for Tennis Frame 3 in Wavelet Domain
(a. Frame 2  b. Reconstructed Frame 3)

5.1 Compression Ratio
To analyze the results of our proposed adaptive prediction, Compression Ratio (CR) and Peak Signal to Noise Ratio (PSNR) parameters are used. Compression Ratio (CR) is defined as the ratio between the number of bits required to store the image before compression (I) and the number of bits required to store the image after compression (O).

Table 1 lists the motion characteristics of each frame over its previous frame, its prediction mode and compression ratio for the Tennis frame sequences (1-9). The prediction threshold is fixed as 0.99. The frames with correlation coefficient 0.99 and above are considered as having, low motion features and their prediction domain is wavelet. The Frames that have correlation coefficient below 0.99 are high motion frames and hence their prediction is direct spatial domain.

Table 1: Compression Ratio For Test Video Sequence (Tennis Frames 1-9)

<table>
<thead>
<tr>
<th>Frame</th>
<th>Correlation coefficient</th>
<th>Prediction domain</th>
<th>Compression ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.983613</td>
<td>Spatial</td>
<td>6.86</td>
</tr>
<tr>
<td>2</td>
<td>0.985742</td>
<td>Spatial</td>
<td>6.73</td>
</tr>
<tr>
<td>3</td>
<td>0.994431</td>
<td>Wavelet</td>
<td>7.91</td>
</tr>
<tr>
<td>4</td>
<td>0.997359</td>
<td>Wavelet</td>
<td>7.25</td>
</tr>
<tr>
<td>5</td>
<td>0.959662</td>
<td>Wavelet</td>
<td>7.93</td>
</tr>
<tr>
<td>6</td>
<td>0.992310</td>
<td>Wavelet</td>
<td>7.96</td>
</tr>
<tr>
<td>7</td>
<td>0.971542</td>
<td>Spatial</td>
<td>6.85</td>
</tr>
<tr>
<td>8</td>
<td>0.981276</td>
<td>Spatial</td>
<td>6.62</td>
</tr>
<tr>
<td>9</td>
<td>0.975463</td>
<td>Spatial</td>
<td>6.37</td>
</tr>
</tbody>
</table>

From the results of table 1 it is obvious that, temporal prediction in wavelet domain has high compression ratio than spatial prediction. However, spatial prediction is faster and motion estimation in the wavelet domain is inefficient for high motion frames, due to shift invariant properties of wavelet transform.

5.2 Peak Signal to Noise Ratio
To analyze the quality of proposed adaptive prediction method, the Peak Signal to Noise Ratio (PSNR) is calculated between the original frame and reconstructed frames by,

\[
PSNR = 10 \log_{10} \left( \frac{255^2}{\text{mse}} \right)
\]  

(9)
where, Mean Square Error (mse) is

$$mse = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{i,j} - x_{i,j})^2$$  \hspace{1cm} (10)

In (10), m and n denote respective number of rows and columns in the image, $y_{i,j}$ is the decompressed image at location (i,j) and $x_{i,j}$ is original image at location (i, j). Table 2 shows the possible combination of spatial-wavelet domains for motion estimation in case of low motion frames and frames with high motion characteristics.

### Table 2: Combination of Domains For Prediction

<table>
<thead>
<tr>
<th>Possible Combination</th>
<th>Low motion frame</th>
<th>High motion frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Spatial</td>
<td>Spatial</td>
</tr>
<tr>
<td>C2</td>
<td>Wavelet</td>
<td>Wavelet</td>
</tr>
<tr>
<td>C3</td>
<td>Wavelet</td>
<td>Spatial</td>
</tr>
</tbody>
</table>

Table 3 gives the performance comparison of the quality parameter in terms of PSNR for the proposed adaptive method with the existing spatial alone and wavelet alone approaches. The PSNR values in C3 column of our various samples of input in Table 3, shows that the temporal prediction using the proposed adaptive method is robust and it can be well adopted for lossless video sequence compression.

### 5.3 Speed Up of Parallel Implementation

The table 4 shows the execution time, of a cluster of 4 workstations and one master server and Table 5 is the speedup of parallel implementation of the temporal predictor. In this paper, only the calculation power of each processor has been taken into account and the communication cost due to message passing is not considered.

### Table 4: Fixed Frames Distribution For Each Slave And Its Execution Time (In Seconds)

<table>
<thead>
<tr>
<th>Slave No.</th>
<th>Frames Processed</th>
<th>Total Time (Sec)</th>
<th>Average Time (Sec)</th>
<th>Frame Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slave #1</td>
<td>12</td>
<td>763.44</td>
<td>63.62</td>
<td>0.94</td>
</tr>
<tr>
<td>Slave #2</td>
<td>12</td>
<td>565.68</td>
<td>47.14</td>
<td>1.27</td>
</tr>
<tr>
<td>Slave #3</td>
<td>12</td>
<td>604.20</td>
<td>50.35</td>
<td>1.19</td>
</tr>
<tr>
<td>Slave #4</td>
<td>12</td>
<td>829.44</td>
<td>69.12</td>
<td>0.86</td>
</tr>
</tbody>
</table>

The speed up for the proposed parallel method is calculated by

$$Speedup = \frac{\text{Execution time}_{\text{sequential}}}{\text{Execution time}_{\text{parallel}}}$$  \hspace{1cm} (11)

The total time taken by a single PC with the said configurations to complete the motion vector estimation and approximation for 48 frames and the maximum time consumed by parallelization of the same operation are listed in Table 5.

### Table 5: Speedup of Parallel Implementation

<table>
<thead>
<tr>
<th>Slave Number</th>
<th>Sequential time (Sec)</th>
<th>Parallel time (Sec)</th>
<th>Speed up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slave #1</td>
<td>2943.36</td>
<td>829.44</td>
<td>3.54</td>
</tr>
<tr>
<td>Slave #2</td>
<td>2202.72</td>
<td>829.44</td>
<td>2.65</td>
</tr>
<tr>
<td>Slave #3</td>
<td>2357.46</td>
<td>829.44</td>
<td>2.84</td>
</tr>
<tr>
<td>Slave #4</td>
<td>2322.80</td>
<td>829.44</td>
<td>3.89</td>
</tr>
</tbody>
</table>

The total time taken by a single PC (slave #3 without communication overhead) to complete the motion vector estimation and approximation for 48 frames is 3053.76 seconds and the time consumed by parallelization the same operation with 2 to 4 slaves and speed up are listed in Table 6.
Table 6: Speedup Of Parallel Implementation

<table>
<thead>
<tr>
<th>No. of Slaves</th>
<th>Total time in Sec</th>
<th>Speed-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1706.71</td>
<td>1.79</td>
</tr>
<tr>
<td>3</td>
<td>1131.02</td>
<td>2.70</td>
</tr>
<tr>
<td>4</td>
<td>850.62</td>
<td>3.59</td>
</tr>
</tbody>
</table>

Figure 9: Speed Up Comparison

Figure 9 shows the speed up curve. We observe from the figure that the speedup for parallel video compression is nearly linear with the number of processors. However, for very large number of processors in the cluster, the overall execution time increases, because of communication overhead. Hence, if communication overhead becomes greater than the computing time, the number of processing nodes in the cluster is too limited.

5.4 Results of Load Balancing

When working with cluster of heterogeneous nodes, the load balancing of an application has a direct impact on the speedup to be achieved as well as in the performance of the parallel system. Hence, predictive load balancing formula as we have defined in (8) is applied to calculate the number of frames to be processed by each node. The data of table 1 shows the frames processed by each slave produced by the algorithm for balanced load distributions.

Table 6: Frames Distribution For Each Slave And Execution Time (In Seconds) As Per Balanced Load Algorithm

<table>
<thead>
<tr>
<th>Slave No.</th>
<th>Frames Processed</th>
<th>Total Time</th>
<th>Average Time</th>
<th>Frame Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slave #1</td>
<td>11</td>
<td>578.82</td>
<td>52.62</td>
<td>1.14</td>
</tr>
<tr>
<td>Slave #2</td>
<td>15</td>
<td>714.60</td>
<td>47.64</td>
<td>1.24</td>
</tr>
<tr>
<td>Slave #3</td>
<td>13</td>
<td>674.18</td>
<td>51.86</td>
<td>1.15</td>
</tr>
<tr>
<td>Slave #4</td>
<td>9</td>
<td>541.71</td>
<td>60.19</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 10 shows the computational load distribution of parallel adaptive temporal prediction algorithm for Tennis sequence. The parallel execution time will be greatly affected by the longest processing time among processors where the workload is high. To cope up with this behavior, the computational load is balanced all over the nodes according to its computing power.

Figure 10: Balanced Load Distributions For Each Slave

The speedup for the proposed load balanced parallel method is calculated by

Throughput = Execution time\(_{\text{fixed}}\) / Execution time\(_{\text{Balanced}}\)

Table 7: Speedup Of Parallel Implementation with Balanced Load
Table 7, shows the computing cost of fixed and balanced load and figure gives the comparison of computing cost of fixed load and balanced load frame allocation among slaves.

![Computing cost of Fixed and Balanced load](image)

**Figure 11: Comparison of Computing Cost Of Fixed and Balanced Load**

From the results obtained, it is observed that, the speedup achieved with the load balancing is slightly increases than fixed load method. This parallel implementation of adaptive temporal prediction for video compression with load balancing on a cluster environment of 1 master and 4 heterogeneous slaves is 1.10 times faster than static load distribution. In real world cluster environment, as the number of slaves increases the speed up will also gets increased, and so notable computing time could be saved with balanced load.

However, Communication between two processing steps is expensive in computing time. This is due to the large amount of regular communications required by message passing interface techniques [4],[10]. Our future work will concentrate on optimal data organization for reducing the communication overhead and developing a component based, distributed motion estimation method for web applications.

6. CONCLUSION

In this paper, we have implemented a parallel temporal prediction model for fast video sequence compression. The motion estimation is done adaptively in spatial domain or wavelet domain, according to its motion features. Hence a good compression rate is achieved for video compression than the existing single domain prediction methods. Further, parallel execution is implemented using client-server architecture on a cluster of local work stations. It is a costless method than other parallel schemes, since computer network is common every where. The speed up comparison shows that, the execution time of video compression can be much reduced by performing the compression of different frames simultaneously, on multiple PCs of a cluster environment. Moreover, an unbalanced workload across the processors can significantly reduces the performance of the parallel program. In order to overcome this issue, a balanced load distribution algorithm based on computing power of each node is introduced. This improvement enhances the speed up and has better behavior than direct fixed load frame distribution. From the experiments conducted and results obtained, this method is found to be very useful for real time on-line applications such as telemedicine, video conferencing and video surveillance system.

REFERENCES


Parallel Adaptive Temporal Prediction with Load Balancing for Fast Video Compression


Author’s Biography

S. Jeyakumar is working as Assistant Professor in Information Technology at Dr. Sivanthi Aditanar College of Engineering, Tiruchendur, India. He has completed BE degree in CSE and MTech in Information Technology. He is presently doing PhD programme in Anna University, Chennai. His research area includes video compression, optical image processing and parallel image processing.
Dr. S. Sundaravadivelu is presently working as Professor in ECE at SSN College of Engineering, Chennai, India. Earlier, he worked at Thiagarajar College of Engineering, Madurai as Head of ECE department. He completed his ME degree from Regional Engineering College, Trichy and awarded PhD by Madurai Kamaraj University. He has over 25 years of teaching and research experience. He is guiding several research scholars. His research interest includes optical signal processing, digital image processing and parallel computing.
Abstract
This paper presents a review of promising techniques for very low bit-rate, below 64 kb/s video coding. Video coding at such low rates will be a crucial technique in visual services, e.g., visual information transmission and storage. A typical application is to transmit moving videophone scenes through the existing analog telephone lines or via a mobile phone. Video compression techniques have to deal with data enriched by one more component which is the temporal coordinate. Compression techniques developed for still images can be either generalized for 3D signals, or a hybrid approach can be defined based on motion compensation. The video compression techniques can be classified into the following four classes: waveform, object based, model based and fractal coding based techniques. The aim of this paper is to provide an overview of these approaches as applied to low bit rate video coding algorithms.

Keywords: SAMCOW, MP-Matching Pursuit, DSC-Distributed Source Coding, MR-Moving Region, MB-Macro Block.

1. Introduction
The importance of visual communications has increased tremendously in the last decades. By very low bit rate video, we mean a compressed data which does not exceed 64 kb/s in the bitrate of its visual portion. The uncompressed source materials [2,3] for this kind of applications typically contain about 25 k pixels in every luminance picture and a quarter of this for the chrominance. The frame rate is also typically about 30 progressive frames per second. This results in an uncompressed bit stream of up to 10 Mb/s [3]. The desired compression ratios in this case are from 150: 1 to 1000: 1. This is a very difficult task because of the relatively small size of the original pictures in the source data.

Rate scalable video compression is appealing for low bit rate applications, such as video telephony and wireless communication, where bandwidth available to an application cannot be guaranteed. The advent of the MPEG standards [23] has fueled the interest in this area due to their rich set of algorithms and tools for compression of various video formats and sizes at different data rates, for content-based manipulation, and for interactive access to audio-visual data.

A number of standards have been defined for the compression of visual information. The MPEG (Moving Picture Expert Group) standards address the compression of video signals. MPEG-1 operates at bit rates of about 1.5 Mbit/s and targets storage and transmission over communication channels as the integrated services digital network (ISDN) or the local area network (LAN). MPEG-2 operates at bit rates around 10 Mbit/s and is designed for the compression of higher resolution video signals. The recommendation H.261 (also known under the acronym px64) was proposed by the International
Telegraph and Telephone Consultative Committee (CCITT, now known as ITU-T). Based on this standard, videoconferencing at bit rates down to 64 kbit/s has become feasible. This requires the capacity of one channel of the ISDN. In the near future, modern visual communications applications will be possible through Public Switched Telephone Networks (PSTN) or mobile networks.

Intensive research has been performed in the last decade to attain this objective [3]. Variations of the recommendation H.261 for very low bit rate applications have been defined as simulation models. For these simulation models, severe blocking artifacts occur at very low data rates. H.263 [1] has emerged to a high compression standard for moving images, not exclusively focusing on very low bit rates. The improvements in H.263 compared to H.261 are mainly obtained by improvements to the motion compensation (MC) scheme.

Much ongoing research is devoted to develop differing drastically from the existing standards by higher bit rates.

1.1 High Compression Image and Video Coding Approaches

High compression image coding has triggered strong interests in recent years. In this type of coding, visible distortions of the original image are accepted in order to obtain very high compression factors. High compression image coders can be split into three distinct groups. The first group is called waveform coding and consists of transform and subband coding. The second group called second generation techniques consists of techniques attempting to describe an image in terms visually meaningful primitives (contour and texture, for example). The third group is object based coding. Fourth one is based on the fractal theory.

A waveform based coding system involves the following steps:
1. Decomposition/transform of the image,
2. Quantization of the transform Coefficients,
3. Source coding of the quantized coefficients.

The first step transforms the image into another representation, where most of the energy is compacted in a few coefficients. The quantization reduces the dynamic range of the coefficient sand the source coding (Encoding) leads to an efficient transmission At compression factors of about 30 to 40 DCT based technique used in most of the compression standards produces blocking artifacts. All the transform coders suffer from this distortion. Unfortunately, the human eye is very sensitive to such a distortion and therefore, block coders are not appropriate for low bit rate image coding. The main artifact at high compression factors (around 50) is due to the Gibbs phenomenon of linear filters and is called ringing effect [3].

In subband coding, an image is split into a set of subband images by using a group of bandpass filters followed by critical subsampling. For low bit rate subband coding (higher than 50) it is of major importance to exploit the existing zero correlation across the sub bands as proposed in [4] in order to maintain a good quality.

The second group of methods is based on second generation techniques. They attempt to decompose the data into visual primitives such as contours and textures. One approach is to divide the image into directional primitives as proposed in [5]. Segmentation based coding techniques [6] extract regions from the image data which are represented by their shape and their textural content.
Three dimensional subbands coding of video has been first introduced by Karlsson et al. [12], the standard subband filters are used for the spatial directions while a DCT derived filter bank is applied to the temporal dimension. The drawback of 3-D subband coding is that the temporal filtering is not performed along the direction of motion. A solution for this is the combination of the temporal SBC component with motion compensation is proposed.

2.1 Motion Estimation

Motion estimation plays an important role in motion compensated video compression, because of its ability to exploit high temporal correlation between successive frames of an image sequence. Although many types of motion estimation algorithm have been developed, the simplicity of the block-matching technique has made it a natural choice for most video compression standards, including MPEG, H.261, and H.263.

**Block Matching Algorithm:** The approach adopted in block-matching algorithms is first to divide each frame into blocks, typically 16x16 pixels. A motion vector is then calculated for each block in the current frame by searching for the best matching block within a limited search area of the previous frame. Compression is achieved by using this best-matched block, indicated by the motion vector, as the predictor for the current block. Various types of the block-matching techniques are developed so far. The full search (FS) method provides the optimal solution by exhaustively evaluating all the possible candidate blocks within the search range in the reference frame. However, massive computation is required in the implementation of FS. In order to speed up the process by reduction in the number of search locations, many fast algorithms have been developed,
such as three-step search (TSS) algorithm and the new three-step search (NTSS) algorithm [13].

Recent studies show that the motion-vector distribution of a real-world image sequence, within the search window, is highly centre biased. Hence, an improved version of the well-known TSS method, the improved three-step search (ITSS) algorithm [13], specifically aiming towards low bit-rate video coding is proposed. The ITSS has much better performance and faster speed than the TSS, and NTSS.

2.2 Motion Parameters and Motion Model

Conventional motion estimators estimate global motions (camera motion) based on the brightness constancy assumption. However, this is not always true as objects may vary in brightness level, depending on the illumination conditions.

Hoi Kok Cheung et.al [14], proposed a new block-based motion estimator which is capable of estimating motion accurately under varying illumination conditions. With the static sprite coding scheme, the static nature of the sprite image does not allow a coherent representation of the variability of the lighting condition in different frames. The local motion estimator in the proposed technique gives accurate motion estimation under varying illumination conditions. Experimental results proved that system also can effectively handle pixel differences caused by brightness change in the spatial and time domains during the sprite construction and frame reconstruction stages.

2.3 Wavelet Based Rate Scalable Video Coding

Wavelet based motion compensated video coders are recently developed one. These coders provide both temporal and spatial scalability. Continuous rate scalable applications can prove valuable in scenarios where the channel is unable to provide a constant bandwidth. Rather than terminating the session, a decoder can adjust the bit rate to use the limited resources, yet produce video of acceptable quality. Such decoders are particularly attractive because of their flexibility.

SAMCOW: A technique known as Scalable Adaptive Motion Compensated Wavelet (SAMCOW) compression SAMCOW is well suited for data rates less than 32 kbps. Several techniques to improve the subjective picture quality of SAMCOW are investigated in paper [15]. These include the use of B frames, half-pixel accuracy, and unrestricted motion vectors. Eduardo Asbun [15] made improvements to the SAMCOW algorithm, including the use of MPEG-type B frames, half-pixel accuracy and unrestricted motion vectors adapted to the particular scalability requirements. The performance of SAMCOW is be comparable that of H.263+.

SAMCOW uses an embedded rate scalable coding strategy to obtain continuous rate scalability. Its main features are adaptive block-based motion compensation in the spatial domain to reduce temporal redundancy and improve image quality at low data rates, and a modified zero tree wavelet image compression schemes used on the intra coded and predictive difference frames that exploits the interdependence between the color components. This codec was implemented using wavelets.

2.4 Multiple Block Pursuit Algorithm

MATCHING PURSUIT (MP), which is a frame-based algorithm, and is a promising method for low bit rate video coding [16]. An MP-based codec yields a better PSNR and perceptual quality than a transform-based codec, and its decoder is simpler [17]. However, it cannot be used in applications that require real time bidirectional communications, because the encoder consumes a
massive amount of computational time. An MP encoder does not obtain all the coefficients in one step, but iteratively finds the frame coefficient that has the largest absolute inner product value between a residual and all the bases. This difficulty prevents the MP algorithm achieving its best performance and its encoder is also complex one.

In MP algorithm, an atom is composed of a base and its corresponding inner product value. The most popular approach for finding an atom is that, a residual frame is divided into blocks and, at each iteration, an atom is found within the block with the highest energy. Although the performance can be improved by finding an atom from more than one block, there is still the issue of the massive number of inner products between a residual and the bases in the blocks. Jain-Liang et.al.[18] proposed a residual in a subspace, spanned by a small number of bases within a few blocks. The bases and the blocks are selected according to the content of the residual, while the coding performance and efficiency are determined by the number of bases and the number of blocks. This algorithm achieves better subjective and objective performances and requires less runtime than one-block algorithms for various sequences at low bit rates.

3. Model Based Coding Distributed Video Coding Techniques

In this type of coding the redundancy between video data collected by cameras are considered in addition to the inherent temporal and spatial redundancy within the video sequence. This consideration significantly reduces the bandwidth.

3.1 Distributed Source Coding

The fundamental ingredient of DSC is binning- a many-to-one mapping of the actual data taken from the sources to a limited number of values. Through binning, the correlation between the sources can be exploited without any communication between the sensors. For two maximally correlated pair of feature pixels from each view, the distributed scalar quantization of the pixel values is used.

In [19] developed a novel multi-terminal, model-based video coding algorithm combining distributed source coding (DSC) and computer vision techniques. The novelty of this work lies in the use of computer vision techniques to reduce inter-camera redundancy in a multicamera setting. Distributed video coding (DSC) is utilized either for the exploitation of temporal correlation in a single video stream, or for better error resilience. Another recent work [20] developed a distributed image coding technique for a multi-camera setting with several restrictive constraints: cameras are located along a horizontal line, the objects are within a certain known range from the cameras, and the image intensity field is piecewise. This paper shows [20] how general framework can be extended to the case of model-based compression. The use of a 3D model reduces the inter-sensor exchange and it suffices to transmit this information only when there is an appearance or disappearance of features. In [20] scheme relies on 3D model-based tracking algorithm that operates independently on each of the video sequences. The tracked features points are coded using a combination of distributed compression and predictive coding schemes.

3.2 Foveation Techniques

The real-time foveation techniques for low bit rate video coding is currently developed one. Foveation is a layer of HVS modeling that describes its inability to perceive an entire visual stimulus at full resolution because of the non-uniform spacing of sensor neurons. This limitation enables the removal of extraneous resolution information
to obtain an increase in compression gain without sacrificing perceived quality. Video coding that incorporates foveation modeling is called foveated video coding. Foveated video coding can provide a significant increase in compression gain beyond the abilities of uniform resolution coder [21]. Apart from coding gain, the computational complexity of a compression algorithm plays a vital role in determining its feasibility. Foveation requires extra processing at the encoder. Although fast foveation techniques have been explored previously [22], the need to combine foveation processing with standard-compliant video coding techniques for real-time operation, requires further research into reducing the complexity overhead. The techniques presented [22] can be implemented using the baseline modes in the video coding standards and do not require any modification, or post processing in the decoder.

4. OBJECT BASED CODING

The subjective quality of reconstructed images may be bad at low bit rate in block based video coding when comparing to the object based coding. Object-based video sequence coding has been intensely investigated in the last few years and is also supported by the new MPEG-4 standard [23]. It has an important advantage over the block-based coding: it allows manipulation of image objects without complete decoding of the stream, and then improves the coding quality and reduces the bit rate. In such a scheme, a prior segmentation map (alpha plane) [23] of the image, which segments the image into objects is known in advance. The object-based approach has been considered as a very promising alternative to the block-based approach. It alleviates the problem of annoying coding effects, such as blocking artifacts and mosquito effects compared to block-based approach at low bit rate, especially when the blocks coincide with boundaries of different objects. The object-based approach can also provide more natural representations of the scene and has another potential benefit of acquiring the depth information of semantically meaningful objects.

4.1 Dynamic Coding

In Marc Chaumont et al [24] proposed a video-object based coding scheme using dynamic coding. The principle of dynamic coding is to set on a competition different coders of on each video object. In [24], a video-object based dynamic coding scheme using four completely different coders of a 3DModel based coder, a Sprite coder, a 2D+t Wavelet coder and an H264/AVC object based coder. The work firstly comprise a global rate-distortion optimization enabling an optimal selection of a coder and its parameters for each object, and secondly the definition of a distortion metric. When comparing traditional H264/AVC (i.e full frame) and the video-object based dynamic coding scheme for bit-rate around 100Kb/s, results are better for the object approach in terms of the PSNR text metric and the visual reconstruction. At very low bit-rates, the video-object based dynamic coding scheme can perform better than H264/AVC. However, at upper bit-rate (more than 250 Kb/s for a CIF 15Hz sequence), gains obtained by using an object approach are not strong enough to compensate object overhead (shape, texture and description overhead).

4.2 Object Coding For Selective Quality Video Communication

A more efficient compressed representation is possible if selective quality can be applied to the objects that compose the scene. For example, in the video telephony application, the quality of the foreground object can be enhanced compared to the background. The background
can be coded, and therefore rendered, at a lower spatial/temporal resolution or signal-to-noise ratio (SNR). Also, faces are essential for visual communication and could therefore be coded at a higher quality. In addition, object-based representation allows greater error protection for essential objects, such as faces and foreground, while transmitting video over noisy channels.

The object based video coding standard MPEG-4 provides functionalities, namely object scalability, error resilience, and coding efficiency. However, the standard itself does not indicate how to obtain these objects. Therefore, the key to utilize these functionalities in MPEG-4 lies in the ability to automatically segment the video sequence in real-time to obtain constituent objects. If multiple views of the same scene are available, then the difference in the views can be used to compute possible delineation between objects in the captured images [25,26]. Alternatively, if only one view is available, specific objects such as human faces can be located and segmented based on some form of computer modeling. In addition, researchers have had some success in using primarily motion, but also color and texture cues, to segment objects in a single view sequence [27, 28].

When two views of a scene are available, a disparity map can be computed from those views and the scene can be segmented into background and foreground objects. As opposed to traditional disparity algorithms, these algorithms compute disparity values by matching blocks [28], image features and multiscale contours. The disparity values are calculated here to segment the objects. In addition, the segmented objects should not contain any holes, even if there is little texture inside the objects. A fast algorithm is required to achieve real-time performance.

4.3 Spatiotemporal Frequency Domain Methods

Frequency or spatiotemporal-frequency domain methods have been developed to complement or overcome some difficulties of spatial-only methods. These approaches are based on the phase shift introduced to the frequency domain representation by spatial translations. They process the entire frames simultaneously, so they are inherently robust to local inaccuracies, like local occlusion, illumination changes and motion discontinuities. Wavelet domain methods [29] are able to localize motions extracted from phase information, but they suffer from inaccuracies at motion discontinuities. The spatiotemporal filtering approaches are able to successfully estimate and localize translations, but they have a high computational cost, as they require the application of many different velocity-tuned filters to the video, in order to give accurate estimates.

In order to address these limitations, a novel approach that combines the results of frequency and spatial domain processing is proposed [30]. In the multiple objects independent translational and rotational motions of the video sequence is analyzed through a combination of spatial- and frequency-domain representations. A novel algorithm is presented [38] for the simultaneous extraction of all objects undergoing translation and the background via a least squares technique that takes place entirely in the Fourier domain. Spatial information is combined with the frequency domain object extraction results, to further refine them. This combined analysis takes advantage of the strengths of both representations, by providing reliable and computationally efficient motion estimates and object segmentation. This algorithm is being a robust to local noise and occlusion, because of its global nature.
4.4 Pattern-Based Video Coding Techniques

The very low bit rate (VLBR) video compression standards, such as H.264, MPEG-4 standards are still however unable to encode moving objects within a 16x16 pixels macro block during motion estimation and compensation, resulting in all 256 residual error values being transmitted for motion compensation regardless of their mobility. VLBR video coding using patterns to represent moving regions in macro blocks has already proved its potentiality over coding standards [31]. The most important applications of these techniques are real time video conferencing, video telephony, and video on demand by mobile terminals. However, given the importance of arbitrary shape patterns for the simplified object segmentation, these techniques have significant potential for commercialization such as the possibility of patenting the arbitrary shaped patterns for the most popular and widely used video coding standard H.264. This algorithm focused on the moving regions of the MBs, through the use of a set of regular 64-pixel pattern templates, from a codebook of patterns in Fig1. If in using some similarity measure, the MR of an MB is well covered by a particular pattern, then the MB can be coded by considering only the 64 pixels of that pattern with the remaining 192 pixels being skipped as

![Pattern Templates](image1.png)

Figure 1: PC Of 32 Regular Shaped, 64 Pixel Patterns, Defined In 16x16 Blocks, Where The White Region Represents 1 (Motion) And The Black Region Represents 0 (No Motion).

Successful pattern matching theoretically has a maximum compression ratio of 4:1 for any MB. The actual achievable compression ratio will be lower due to the computing overheads for handling an additional MB type, the pattern identification numbering and pattern matching errors.

In [32], a real time pattern selection (RTPS) algorithm, which uses a pattern relevance and similarity metric to achieve faster pattern selection from a large codebook, is proposed. For each applicable macro block, the relevance metric is applied to create a customized pattern codebook (CPC) from which the best pattern is selected using the similarity metric. The CPC size is adapted to facilitate real-time selection.

The RTPS algorithm [32] the size of the CPC is controlled within predefined bounds, to adapt the computational complexity of the pattern selection process, so ensuring real time operation. RTPS is able to process arbitrary-sized codebooks while this real-time constraint is upheld. Furthermore, the computational overhead of the similarity metric is reduced significantly by performing the processing on a quadrant-by-quadrant basis with the option to terminate whenever the measure exceeds a predefined threshold value. Any pattern selection algorithm using the same set of patterns for a video sequence is termed as fixed algorithm. The eight-pattern algorithm is, therefore, referred to as the Fixed-8 algorithm. Overall, the computational efficiency for the RTPS algorithm is superior to both the Fixed-8 algorithm and H.263 low bit rate video coding standard.

5. Fractal Based Coding

The introduction of a fractal coding scheme turns out advantageous especially at very low bit rates (8–64 kbit/s). The fractal coding concept originates from
Barnsley’s idea to exploit self similar structures in real world images for compression purposes [33]. A first practical implementation capable of encoding grey scale images has been proposed by Jacquin [34]. Several improvements and modifications e.g. [34, 35] have been reported since then, but the basic concept of block wise approximation of the entire image by parts of itself remained the same.

According to Banach’s fixed point theorem the sequence of reconstructed images converges for any arbitrary initial image to the fixed point of the transformation which is the original image. Compression is achieved if the transformation parameters can be described more compactly than the original image.

Former investigations performed by researchers indicate that fractal coding schemes are suited for applications with a demand for extremely high compression ratios. Due to the increased encoding complexity for visual loss-free coding at low compression ratios, fractal coding schemes cannot play of its advantages compared with advanced subband or transform coders. But it is more suitable for low rate video coding with applications in mobile video telephony, teleconferencing and narrow band ISDN distributed audio-visual services which rise from 4.8 to 64 kbit/s because of its high coding gain.

An efficient scheme for video coding is presented which utilizes progressive fractal coding, called wavelet based fractal approximation (WBFA) [36] and motion compensation (MC). In the scheme, the MC error frames are encoded by the progressive fractal coding. Considering the severe localization of the MC errors, irregular sampling of domain blocks in a fine step is used instead of regular sampling of domain blocks in a coarse step. It is shown that the fractal video coder presented is competitive with or superior to H.263. It also has a relatively simple structure and reveals the characteristics of non-iterative decoding, and relatively fast encoding and decoding. It is found to be more suitable for low bit rate video coding.

A novel object-based fractal monocular and stereo video compression scheme with quad tree-based motion and disparity compensation is proposed in [37]. In [37], Fractal coding is adopted and each object is encoded independently by a prior image segmentation alpha plane, which is defined as in MPEG-4. The performance results shows that it is more suitable for low bit rate coding.

6. Conclusion

The four approaches as, waveform based, model based, object based and fractal based video compression for very low bit rate video coding are discussed in this paper. The performance of these algorithms is compared as in the Table1.

The performance of these algorithms can be improved in future as follows.

(i) Adaptation of the parameters in MP Techniques such as number of blocks and number of bases, based on the input video sequence [18].

(ii) The PSNR and computation efficiency of the pattern based coding techniques can be improved by analyzing pattern similarity and relevance measurements [32].

(iii) In [31], Future directions in Spatio-temporal motion estimation and object extraction involve by analyzing the joint use of spatial and frequency data for the more complex motions, including random variations, as well as the examination and analysis of the motion of nonrigid bodies.

(iv) Very low bit rate coding multiview dynamic coding could be developed from the existing single view coding techniques.
Fractal based encoding speed may be improved by analyzing the fractal partition schemes by constraining the dissimilarity measure and the domain block search for single object.

Table 1: Performance Comparison of Low Bit Rate Coding Algorithms

<table>
<thead>
<tr>
<th>Technique</th>
<th>PSNR</th>
<th>Run Time of Encoder</th>
<th>Encoder Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP Update Algorithm (waveform based) [18]</td>
<td>PSNR increases with increase in L: 32.63dB</td>
<td>205.33 sec</td>
<td>Complexity increases with increase in L (base)</td>
</tr>
<tr>
<td>Foveation model based algorithms [22]</td>
<td>Comparable PSNR value with H 263</td>
<td>Encoding time is high.</td>
<td>Preprocessing is required and high complexity</td>
</tr>
<tr>
<td>RTPS Algorithm (Object based) [32]</td>
<td>1.52dB higher than H 263 (37.11dB)</td>
<td>200.00 sec</td>
<td>Less Computation complexity</td>
</tr>
<tr>
<td>Progressive Fractal Video coder (WBFA) [36]</td>
<td>.45 dB higher than H 263</td>
<td>221.25 sec</td>
<td>Less complexity when comparing with conventional fractal based codec</td>
</tr>
</tbody>
</table>

References


[16] Keman Yu, Jiang Li, Jizheng Xu and Shipeng Li, Microsoft Research Asia, Beijing, China, “Very Low bit Rate Watercolor Video”.
[30] Alexia Briassouli, Member, IEEE and Narendra Ahuja, Fellow, IEEE, “Integration of Frequency and Space for Multiple Motion Estimation and Shape-Independent Object Segmentation”, IEEE


[36] Bernd Hurtgen and Peter Buttgen, “Fractal approach to low rate video coding”, Visual Communication and Image Processing, SPIE/VisCom’93, Nov.’93, Boston, MA, USA.


Author’s Biography

M. Thamarai received the M.E. degree in Applied Electronics from the Anna University Chennai in 2004. She is currently pursuing the Ph.D. degree at Government college of Technology, Coimbatore under Anna University Chennai as a part time Scholar. She has been working as an Assistant professor in Karpagam College of Engineering, Coimbatore since 2005. Her research interests are Digital image processing and low bit rate video coding.

R. Shanmugalakshmi was born in Coimbatore, India. She received the Master of Engineering degree in year 1990 and Ph.D in the year 2005 from Bharathiar University, Coimbatore. She is working as Assistant Professor in the department of Computer science and Engineering, Government College of Technology, Coimbatore. Her research interest includes Image compression, Genetic Algorithms and neural Networks. She has published more than 30 papers in National and International Journals.
Index Based Internet Traffic Sharing Analysis of Users by a Markov Chain Probability Model

D. Shukla
Dept.of Mathematics and Statistics, Dr.H.S. Gour University, Sagar, M.P – 470003. Email: diwakarshukla@rediffmail.com

Sanjay Thakur
Dept.of Computer Science and Applications, Dr.H.S.Gour University, Sagar, M.P – 470003. Email: sanjaymca2002@yahoo.com

ABSTRACT
The increasing demand of Internet access caused a variety of congestions in the networks, which could be reduced by the better management offered due to service providers. This leads to an assessment of the quality of service (QoS) of operators by the users. This paper takes into account the set-up of two Internet Service Provider (ISP) environments and a Markov chain model is applied over the Internet access behavior of the user indices for different types of users are defined and compared under model approach using simulation study. The blocking probability of the competitor ISP has a significant impact on the initial preference.

Keywords: Markov Chain Model, Initial Preference, Call-by-Call Basis, Internet Service Provider [ISP or operators], Internet Access, Network Congestion, Users Behaviour.

1. INTRODUCTION
In every walk of life, the use of Internet is being in practice in modern days of electronic communication. More and more people are joining to the club of Internet users in the entire world. This has generated a high amount of traffic load on the network due to rigorous emergence of call requests every minute. Excess traffic load constitutes immense trouble in the form of congestion in the flow of information through one or more networks. Large numbers of Internet Service providers (ISP) are coming up in the market, making this a business of quality oriented. Consumers need high quality of service for call connectivity and this very ground used to decide about their initial market preference to an ISP. Some users may be loyal and dedicated to an ISP and some may casual in their approach. The pre- set starting choices (or liking) to an ISP could be an evaluation factor for market capture in a competitive market.

This paper presents an analysis of initial preference of an ISP by the different categories of users present in the market. Index for user categories are formulated and compared with each other. Naldi [16] discussed quality of service (QoS) and user behavior under two-operator environment in the setup of four states only. We add a fifth state whose impact has been examined in this content. Deriving further motivation from Naldi [15],[17] and applying theories, techniques of Medhi [12], Perzen [11], Yuan and Lygeros [3], a Markov chain model is used to study the behavior of various groups of consumers in the competitive market situation. Some other useful contributions for the use of Markov chain models are due to [1], [2], [4], [5], [6], [7], [8], [9], [10], [13], [14] and [18]. In order to avoid the mathematical complications, only two ISP are taken into consideration in this contest.
2. USER BEHAVIOUR FOR INTERNET ACCESS

Suppose ISP1 and ISP2 are two Internet service providers in a market, each having predetermined market capture by a% and b% respectively (a + b = 100%). Assume the followings:-

(i) A user connects his call through either ISP1 or by ISP2.

(ii) User attempts for an ISP only once and then shift to the next ISP in next call attempt and so on.

(iii) There are three other options for a user like (a) go to rest (b) abandon the attempt process (c) get success during call attempt.

(iv) The initial probability of selection to ISP1 is p = a/100, and to ISP2 is (1-p) = b/100.

(v) The blocking probability in network due to congestion experienced by ISP1 is L1 and by ISP2 is L2.

3. MARKOV CHAIN MODEL FOR USER BEHAVIOUR

Let \( Y(n) \), \( n \geq 0 \) be a Markov chain having transitions over the state space \{ISP1, ISP2, RS, SS, AP\}, where \( Y(n) \) denotes the position of random variable Y at the nth call attempt \( (n \geq 0) \), and five states in state space are:

State ISP1: First Internet service provider.

State ISP2: Second Internet service provider.

State RS: Taking rest for a short duration

State SS: Success obtained in call connection

State AP: Leaving the process of call attempt

Suppose the user is on ISP1 at nth attempt. If this call is blocked with the probability L1 then he shifts either to ISP2 or to RS state in the \((n + 1)th\) attempt. User can not be at the same state in two consecutive call attempts except at SS and AP. User abandons the call attempt process at \((n + 1)th\) attempt with probability pA. If he reaches to RS from ISP1 in \(n^th\) attempts then in next \((n + 1)th\) he may either be on ISP1 with probability r, or on ISP2 with probability \(1 - r\). From RS user can not move to states SS and AP. The diagrammatic form of state transition mechanism is in fig 3.1 and transition probability matrix is in fig 3.2.

![Figure 3.1: Transition Diagram](image)

![Figure 3.2: Transition Probability Matrix](image)

(a) Faithful User (FU)

A user who is dedicated to ISP1, otherwise prefer to rest state RS but does not attempt for ISP2. The converse of
it happens for ISP$_2$. A group of this kind of users is defined as faithful users (FU) for ISP$_1$ or ISP$_2$.\\n
(b) Partially Impatient User (PIU)\\nA user who attempts between ISP$_1$ and ISP$_2$ only all the time until call completes or abandon the process but never prefers to reach to RS. A group of this behavior is categorized as PIU.\\n
(c) Completely Impatient User (CIU)\\nA user who attempts to ISP$_1$ or ISP$_2$ or RS or abandon in nth attempt. A group of user with such behavior is termed as CIU.\\n
The nth step transitions probability for FU to ISP$_1$ for $n \geq 0$ is\\n$$P\left[ Y^{(2n)} = ISP_1 \right] = pE^n$$\\n$$P\left[ Y^{(2n+1)} = ISP_1 \right] = 0$$\\n(1)\\n$$P\left[ Y^{(2n)} = ISP_2 \right] = (1-p)D^n$$\\n$$P\left[ Y^{(2n+1)} = ISP_2 \right] = 0$$\\n(2)\\nwhere $E = B_r . D = B_t(1-r) . B_1 = L_1(1-p_A)p_R$, $n = 0,1,2,3 - - -$\\n
For PIU, the nth step transition probability is :-\\n$$P\left[ Y^{(2n)} = ISP_1 \right] = pC^n$$\\n$$P\left[ Y^{(2n+1)} = ISP_1 \right] = (1-p)A_2C^n$$\\n$$P\left[ Y^{(2n)} = ISP_2 \right] = (1-p)C^n$$\\n$$P\left[ Y^{(2n+1)} = ISP_2 \right] = pA_1C^n$$\\n(3)\\nFor Completely Impatient User (CIU), the nth attempt approximate expressions of probability are:\\n$$P\left[ y^{(2n)} = ISP_1 \right] = p(C+D)^n$$\\n$$P\left[ y^{(2n+1)} = ISP_1 \right] = (1-p)A_2(C+D+E)^n$$\\n$$P\left[ y^{(2n)} = ISP_2 \right] = (1-p)(C+D)^n$$\\n$$P\left[ y^{(2n+1)} = ISP_2 \right] = pA_1(C+D+E)^n$$\\n(4)\\n4. User Index\\nWe define user indices for all the three user categories FU, PIU, CIU for operator ISP$_1$ as:\\n(A) For Faithful Users (FU)\\n$$B_1 = \frac{L_1(1-p_A)p_R}{[L_1(1-p_A)p_R + L_2(1-p_R)p_A]}$$\\n(B) For Partially Impatient Users (PIU)\\n$$B_2 = \frac{L_2(1-p_A)p_R}{[L_1(1-p_A)p_R + L_2(1-p_R)p_A]}$$\\n(C) For Completely Impatient Users (CIU)\\n$$B_3 = \frac{L_1(1-p_A)p_R}{[L_1(1-p_A)p_R + L_2(1-p_R)p_A]}$$\\nSimilar expressions of user indices could be written for operator ISP$_2$. These indices are function of four parameters $L_1, L_2, p_A$ and $r$.\n
1541
5. **Simulation Based Index Analysis**

Indices are examined over number of increasing attempts as function of blocking probabilities.

From fig 5.1 – 5.4 (Index of FU for ISP\textsubscript{1} over varying L\textsubscript{2}), it is evident that with the increase of blocking probability of opponent, the operator ISP\textsubscript{1} looses index value for faithful users. But, however, with the increase of rest state probability r, this index has growing trend. So the rest state plays an important role in the improvement of FU. The increase in p\textsubscript{R} value also have a positive impact on growth of index value.

**Figure 5.1** : L\textsubscript{1} = 0.3, p\textsubscript{A} = 0.05, p\textsubscript{R} = 0.03, r = 0.3

**Figure 5.2** : L\textsubscript{1} = 0.3, p\textsubscript{A} = 0.05, p\textsubscript{R} = 0.03, r = 0.7

**Figure 5.3** : L\textsubscript{1} = 0.3, p\textsubscript{A} = 0.05, p\textsubscript{R} = 0.07, r = 0.3

**Figure 5.4** : L\textsubscript{1} = 0.7, p\textsubscript{A} = 0.05, p\textsubscript{R} = 0.03, r = 0.3

**Figure 5.5** : L\textsubscript{1} = 0.3, p\textsubscript{A} = 0.05, p\textsubscript{R} = 0.03, r = 0.3

**Figure 5.6** : L\textsubscript{1} = 0.3, p\textsubscript{A} = 0.05, p\textsubscript{R} = 0.07, r = 0.3

**Figure 5.7** : L\textsubscript{1} = 0.7, p\textsubscript{A} = 0.05, p\textsubscript{R} = 0.03, r = 0.3
Looking over to fig. 5.5 - 5.8 (Index of PIU for ISP₁ over varying $L₂$), the opponent’s blocking probability $L₂$ also has an impact on the group strength of partially impatient users of ISP₁. The index for PIU decreases with increasing $r$ whereas simply with high $L₂$, the index shows almost a stable pattern over increasing number of attempts.

On comparing index of completely impatient user (CIU), (from fig. 5.9-5.12) with increasing $L₂$, the probability index for CIU increases. When $r$ is high, the index increases for the lower values of $L₁$, Variation of $pₐ$ has smaller effect than that of $pₙ$ and $r$. The CIU index gets stable value over larger number of attempts with small $L₁, L₂, pₐ$ and $pₙ$. Overall views indicate the probability index for FU has a constant fluctuating nature over large number of attempts. Contrary to this, index of PIU has little fluctuation preferably in the range (0.3 – 0.5). Moreover stability pattern occurs in this range too over large number of attempts. The smaller choice of parameter is preferable and advisable. The index variation of PIU and CIU are very much similar than FU but the stability pattern over large attempt is shifted to higher value (0.4 – 0.6) than the PIU.
6. Conclusion

With the hike of blocking chance of opponent, the operator ISP_1 looses index of faithful users but with little increment in rest state probability this index increases. The rest state has an important role in the development of major stuff of faithful users. Actual benefit starts when the proportion of faithful users is high for an operator. Opponent’s blocking probability L_2 does not affect much to the index of partially impatient users over increasing number of attempts. However, the rest state probability also affects this index. Coming over to the third category, the increase in L_2 has a positive impact to the index increment for completely impatient users. Overall, this is to observe that blocking probability level has a vital role in causing the variations in user indices. The rest state probability also affects these indices. Therefore, it is recommended to open up a restaurant outside to an Internet shop to act like a rest state which would contribute much in attracting Internet traffic.

References


Index Based Internet Traffic Sharing Analysis of Users by a Markov Chain Probability Model


Author’s Biography

Dr. Diwakar Shukla is working as a Professor in the Department of Mathematics and Statistics, Sagar University, Sagar, M.P., India and having over 19 years experience of teaching to P.G. and U.G. classes. He obtained M.Sc. (Stat.), Ph.D. (Stat.) degrees from Banaras Hindu University, Varanasi, UP and served the Devi Ahilya University, Indore, M.P., India as a Lecturer for nine years and obtained the degree of M. Tech. (Computer Science) from there. During Ph.D., he was awarded junior and senior research fellowships of CSIR, New Delhi, through Fellowship Examination (NET) of 1983. Till now, he has published more than 50 research papers in national and international journals and participated in more that 35 seminars / conferences at national level. He is the recipient of MPCOST Young Scientist Award, ISAS Young Scientist Medal, UGC Career Award and UGC visiting fellow to Amerawati University, Maharashtra. He also served as a Professor to the Lucknow University, Lucknow, U.P., for one year and visited abroad to Sydney (Australia) and Shanghai (China) for the conference participation. He has supervised seven Ph.D. theses in Statistics and Computer Science both; and eight research scholars presently enrolled for their doctoral degree under his guidance. He is a member of 10 learned bodies of Statistics and Computer Science. The area of research he works for are: Sampling theory, Graph theory, Stochastic modeling, Computer networks, CPU scheduling, Operating systems and Operation research.

Sanjay Thakur has completed M.C.A. degree from H.S. Gour University, Sagar, MP, India in 2002. He is presently working as a Lecturer in the Department of Computer Science & Applications in the same University since 2007. He did his research in the field of Computer Networks. In this field, he has the authored and co-authored 10 research papers in National/International Journals and proceedings. His current research interest is to the area of Stochastic Modeling of Switching system, Computer networks and Internet traffic sharing analysis.
Bionet: ART1 Neural Network Model for HIV/AIDS Decision Support System

M. Lilly Florence

P. Balasubramanias

ABSTRACT
Since the detection of HIV infection, the infected number is growing very fast in the country. Even though in medical field has many decision making system there is no appropriate clinical system for HIV. This network model designed to identify how long the life of HIV patient can prolong with the calculated ratio of regimens. This model is designed based on ART1 network. To learn this model 300 patient details are taken and trained the network. Among this, particularly some 25 patients are observed and study the improvement in their medical results. There are 10 parameter has taken for this network.

Research in the area of clustering the patients based on their medical history and regimens. There are two clusters that is, the first group belongs those patients who are not able to prolong their life more than of 10 years if they follow certain regimens. The second group of patients they can be alive more than 10 years with certain regimens.

Keywords: ART1 network program implemented in Mat Lab, Data from the famous AIDS Government Hospital, Tamil Nadu.

INTRODUCTION
Since the detection of HIV infection in Tamil Nadu in 1986 infected number is growing very fast in the country. It is reported till Sep. 1999,88,775 seropositive have been reported and seropositive rate has gone up to 25.12 per thousand. The report is from 32 states and UTs in the country.

The natural history of any disease refers to the stages through which a disease passes, in the absence of any intervention. The natural history of HIV infection begins as soon as virus enters the body of a susceptible host through any of the routes of transmission.

Clinical trail system for HIV/AIDS (Human Immunodeficiency Virus / Acquired immunodeficiency Syndrome) is a complex one. This is the case because every patient is unique with his/her own history, set of genetic traits, predisposition to side effects and prognosis. Additionally, many symptoms and diagnoses are inherently imprecise in their definition and difficult to measure. Although clinical trial data provide excellent information regarding expected treatment outcomes for large groups of patients, the prediction of actual treatment outcomes and clinical courses for a particular individual patient may be subject to a considerable degree of uncertainty. With all these uncertainties inherent to the clinical decision process, a clinician's subjective judgment plays a vital role in making sound treatment decisions for individual patients. Various patient specific factors make it difficult to objectively and quantitatively compare various treatment decisions made by different physicians for particular patients. Consequently, inconsistent and sub optimal treatment outcomes can occur even for quite similar patient.
At present a systematic decision making and optimization technology capable of handling all these difficulties, although much desired clinically is still not available. In this research, clustering the HIV patients based on their medical history and the regimens using ART1 network model. The regimens for the patient are based on some factors such as age, weight of the patient, CD4 count, HB rate, and some other parameters. So based on the regimens how long the patient can prolong their life if they continuously taking treatment. Two groups of patients are considering depends on their medical parameter. One group of patients will not have life more than 10 years and another group of patients can have more than 10 years of life. All these groups are considering having the life more than the maximum years unless otherwise they should follow certain regimens. With this model it decides the regimens for each cluster.

**Introduction to ART1**

Adaptive Resonance Theory was developed by Carpenter and Grossberg (Fausett, 1994). One form, ART1, is two fields of units, the F1 units and F2 units, together with a reset unit to control the degree of similarity of patterns placed on the same cluster unit.

**Basic Architecture:**

The basic architecture of adaptive resonance neural net involves three groups of neurons.

- Input processing field – F1 layer
- Cluster units – F2 layer
- Reset mechanism

**Basic Operations**

Learning trail: the activation of all units should be zero and the F2 units are made inactive.

Controlling the degree of similarity: it is controlled by vigilance parameter.

Reset Mechanism states: it control the state of each node in F2 layer. The F2 layer node is present in any one of the three states.

1. **Active – “ON”**
2. **Inactive – “OFF”**
3. **Inhibit – “OFF”**

There are two types of learning fast learning and slow learning. The algorithm used for this model as follows;

1. Initialize parameters
   
   \[ L>1 \text{ and } 0<\eta<1, \]
   
   \[ 0<b_{ij}(0)<L/(L-1+\eta) \text{ , } t_{ij}(0)=1. \]

2. Send input signal from F1(a) to F1(b) layer.
   \[ X_i = S_i \text{ i.e. input vector} \]

3. Compute the norm of vector x
   \[ ||x||=\bullet X_i \]

4. Update the weight
   \[ b_{ij}(\text{new}) = L X_i / (L-1+||X||) \]
   \[ t_{ij}(\text{new})=X_i \text{ repeat the steps with next input vectors.} \]

**Some Clinical Variables Important To HIV/AIDS**

According to current guidelines treatment should focus on achieving the maximum suppression of symptoms for as long as possible. This aggressive approach is known as Highly Active Anti-retroviral Therapy (HAART). The aim of HAART is to reduce the amount of virus in the blood to very low or even non detectable levels although this doesn’t mean the virus is gone. This is usually accomplished with a combination of three or more drugs. But the treatment guidelines also emphasize the importance of quality of life. Thus the goal of AIDS treatment is to find the strongest possible regimens that is also simple and has the fewest side effects.
The treatment for HIV/AIDS patients depends on some important factors such as age, weight, CD4 count, HB, the details of TB etc. The regimens are based on their weight and for patients above 35 years they are taking the constant regimens. Commonly the regimens which are prescribed by the doctors are Zidovudin or Stauvdine, Lamivudine and Nevirapine or Efavirenz. This combination of the regimens varies depending on patients with their medical history. The regimens are based on certain specification. This specification is related with their weight and if the patient is above 35 years, they will follow the constant regimen immaterial of their weight. But in this research all the patient’s regimen specification is based on certain factors. There is no common specification.

Methods
The first issue is the selection of input variables to ANN has a significant impact on the prediction performance of any ANN. In addition to a need to determine the applicability or correlation of input variables to the prediction problem, which should be determined by a medical domain expert each data variable has a direct cost to the patient with respect to time, health risk and patient history. Therefore, the suggested blanket approach of including every possible variable that may affect the dependent variable must be evaluated with respect to the outcomes and life time of patients.

The second issue regards training set size and construction. Since medical data is typically abundant (as long as resources are available to collect data from patient record or other databases) the size of the training sample set should not be concern. The actual implementation of an ANN model will be prospective with respect to the training samples. A common approach used in ANN training for medical domains is to divide the collection of data samples into two training group and testing group.

The third consideration in developing ANN solution to medical problem is distribution of the training sample and test sample. The ANN model must be given an amount of training samples from second group that is comparable in quantity with the first group which will help to optimize the generalization performance of the ANN. In this research 300 patients are taken as training set data and 200 patients’ data are taken as holdout sample set. 14 months of data used for training and 12 months of data used for hold out sample. Three months once the patients are asked to take viral load test and study the performance. Compare the performance with our network. Three technologies measure HIV viral load in the blood-reverse transcription polymerase chain reaction (RT-PCR), branched DNA (BDNA) and nucleic acid sequence based amplification.

The actual input data has been collected from the AIDS centre in a famous Hospital in Tamil Nadu, India. Two sets of patients consider for this research. Total of 500 patients were taken for research. Determining the optimal set of independent variables for predicting the regimens which prove the CD4 count and prolong the life of patients must take into account. The following variables were identified, age, sex, weight, CD4 count, HB rate, TB rate and CD8 and the three types of regimens.

Experimental Results
There are two groups of patients results in this research, one group with particular regimens which depends on their medical history and their life will be maximum of 10 years unless otherwise they should take the treatment regularly. Another group of patient can hold their life more than 10 years if they follow certain regimens.
Result from MATLAB execution for 100 patients:

This is the input vectors for this ART1 model (only 50 patients)

The actual value is decoded with maximum value. The maximum value for each parameter is considered with a help of the famous Doctors who is giving treatment for HIV/AIDS patients.

<table>
<thead>
<tr>
<th>1</th>
<th>1</th>
<th>0</th>
<th>ZLE</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>ZLN</th>
<th>0</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>SLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ZLE</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>ZLN</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>SLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ZLE</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>ZLN</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>ZLE</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1769</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>SLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>ZLN</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ZLE</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>SLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>ZLN</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>ZLE</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1202</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>SLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2598</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>ZLN</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>ZLN</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>SLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3592</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>ZLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>ZLE</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>SLN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1023</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

1549
The output of this input vector is either 0 or 1. If it is 1 it is in active state i.e. the patient can prolong their life more than 10 years. If it is 0 it is Inactive state i.e. the patient can prolong their life maximum of 10 years if they should take follow the corresponding regimens regularly.

The group of patient with 0 outputs will follow the regimens based on their weight and other factors. But the group of patient with 1 output is following the constant regimens immaterial of their medical history. With 6 months observation if they follow the constant regimens there is no improvement in their CD4 count.

CONCLUSION

Acquired immunodeficiency syndrome (AIDS) is a set of symptoms and infections resulting from the damage to the human immune system caused by the human immunodeficiency virus (HIV). HIV infection in human beings transmitted through direct contact of bloodstream with a bodily fluid containing HIV, such as blood, semen, vaginal fluid, preseminal fluid and breast milk and also contaminated needles, from mother to fetus during pregnancy, childbirth. In medical field very difficult to determine history of patients and life time. Because every patient is unique with them own history, set of genetic traits, predisposition to side effect, and prognosis.

Although, some medical parameters such as age, weight of the patient, CD4 count, HB rate are taken for predicting that how long the patient can prolong them life if they are in treatment. In this research, above the data’s could be used with ART1 network program implemented in Matlab. This methodology may address the problems very easily which are arises in medical field while physician treating the patients with HIV. In this method is not a time consuming process compare to others and also with in a short time, it can show the patient living years either it’s >10 or <10 years. 0 or 1 is indicating clearly everything based on the patient medical parameters. Herewith we notify to scientific community it is a novel approach in dynamic environment for finding patients life time due to the development of information technology with less cost. We currently investigating and evaluating this program and the cluster for accepting the medical parameters of other severe pathogen causing diseases.
REFERENCES


**Author’s Biography**

Lilly Florence received her B.Sc. Mathematics at MS University in the year 1995 and also she earned her MCA in the same University. M.Tech., Information Technology at Punjabi University. Currently she is doing her research at Mother Teresa Women’s University. She is working as an Asst. Prof. in the department of MCA at Adhiyamaan College of Engineering, Hosur. Her Current research interests include the bioinformatics, neural networks.

Dr.Balasubramanie, received his Ph.D in computer science and now he is working as a professor in dept of CT (PG) Kongu Engineering College. His current research interests include AI, Neural Network, Network Security, Data Mining, etc.
Generalized Neuron Based Digital Communication Channel Equalizer

Vineeta Choudhary1                               B. K. Joshi2

1Reader, Electronics & Communication Engineering, Ujjain Engineering College, Ujjain, India - 456010. Phone: 0734-2516606 Email: vinita1988@rediffmail.com

2Professor, Electronics & telecommunication Engineering and Computer Engineering, Military College of telecommunication Engineering, Mhow, India. Email: drbkjoshi@gmail.com

Abstract

Equalization is necessary in digital communication system to mitigate the effect of intersymbol interference (ISI) and other nonlinear distortions. In order to reduce complexity the application of generalized neuron (GN) to adaptive channel equalization in a digital communication system with duo-binary signals is investigated. It uses only a single GN thus there is no problem of selection of initial architecture of the neural network giving optimum performance. Low complexity and fast convergence characteristic of GN based equalizer make it suitable for real time application. Bit error rate (BER) over a wide range of signal to noise ratio (SNR) is noted. It has been shown that BER performance approaches to optimal Bayesian solution.

Keywords: Digital communication, ISI, Channel equalizer, ANN, GN.

1. INTRODUCTION

BAND-limited high speed digital transmission suffers from ISI and various other noise sources. Nonlinear distortion is a significant factor hindering further increase in attainable data rate. Equalization is necessary at the receiver to overcome these channel impairments [1]. Since communication channels are time varying in nature hence adaptive equalization is required. Figure 1 shows the simplified model of a discrete time transmission model of a digital communication system. Figure 2 shows the non-linear channel model. NL represents the nonlinearities involved.

$s[k]$, is the original sequence to be transmitted, where $k$ is any time instant. The block channel represents the combined response of the transmitting filter, transmission media and RF/IF sections of the receiver filter. $q[k]$, is the additive white Gaussian noise (AWGN) that corrupts the signal.

Figure 1: Simplified Block Diagram Of Discrete Time Transmission Model

Figure 2: Nonlinear Channel Model
the channel output $y'[k]$. The channel output $y'[k]$ corrupted with AWGN $q[k]$ forms the input to the equalizer $y[k]$. The equalizer at the receiver makes an estimate $s'[k-d]$, of the expected delayed transmitted sequence $s[k-d]$, from the knowledge of channel output $y[k]$, present and past values. Where $d$, is equalizer delay. The difference in the estimated sequence and expected delayed transmitted sequence is minimized during training of the equalizer periodically.

Band-limited communication channels are generally modeled as digital FIR filters represented as

$$H(z) = \sum_{i=0}^{N} h_i z^{-i}$$

(1)

Where, $N$ is the channel order. The channel output vector can be represented as $y(k) = \begin{bmatrix} y(k), y(k-1), \ldots, y(k-m+1) \end{bmatrix}^T$, (2) Where, $m$ is the order of equalizer.

Conventionally, equalization has been considered as a deconvolution problem, where a finite impulse response filter (FIR) based linear transverse equalizer (LTE) is used to invert the channel response and its parameters are adjusted using minimum mean square error (MMSE) criterion [1]. Least mean square error (LMS) algorithm is mostly used to tune the tap weights of the filters to minimize error iteratively. Linear filter based equalizers performance degrades under severe nonlinear distortion conditions. The maximum likelihood sequence estimation (MLSE) [2] requires batch processing of the entire received sequence, gives nearly optimum results but its very high computational complexity restricts its use hence practically not suitable and makes symbol by symbol detection as a good option. The equalization can be treated as a classification problem where, the job of an equalizer is to assign the received signal to one of the signal constellation. The optimal solution (least misclassification) to this classification problem is given by Bay’s theory. The channel input vector for an $m^{th}$ order equalizer is given by

$$s(k) = \begin{bmatrix} s(k), s(k-1), \ldots, s(k-m+1-N) \end{bmatrix}^T$$

(3) And can take $Ns = 2^{*m}$, different values. Different, $Ns$ possible values of noiseless channel output vector is given by

$$y'(k) = \begin{bmatrix} y'(k), y'(k-1), \ldots, y'(k-m+1) \end{bmatrix}^T$$

(4)

Which, is to be divided into two classes

$$Y_+ = \{y'(k) \mid s(k-d) = 1\}$$

$$Y_- = \{y'(k) \mid s(k-d) = -1\}$$

(5)

$y(k)$ is a random process having conditional Gaussian density functions centered at each of the $Y_i$ and $Y_i^-$, where $i \in \{1, 2, \ldots, Ns/2\}$.

If the transmitted sequence $s(k)$ is an independent identically distributed (i. i. d.) and equi-probable binary sequence with values $\{+1, -1\}$. For this sequence the optimal solution to this classification task is given by Bay’s theory as [3]

$$s'(k-d) = \text{sgn}(f_{\theta}(y(k))) = \begin{cases} +1, & f_{\theta}(y(k)) \geq 0 \\ -1, & f_{\theta}(y(k)) < 0 \end{cases}$$

(6)

$$f_{\theta}(k) = \sum_{j \in i^+} \exp \left(-\frac{\|y(k) - Y_i^+\|^2}{2\sigma_n^2} \right)$$

$$- \sum_{j \in i^-} \exp \left(-\frac{\|y(k) - Y_i^-\|^2}{2\sigma_n^2} \right)$$

(7)
Where $\sigma^2_n$ is the variance of the AWGN. Good equalizers for this channel will then approximate the function given by (7) as a decision boundary for classification, which is a nonlinear function. Bayesian equalizer provides the lower performance bound for symbol-by-symbol equalizers in terms of probability of error or BER. Nonlinear mapping capability of artificial neural network (ANN) and fuzzy logic make them a suitable choice for nonlinear equalization. Several equalizers are developed in the past to address this problem using ANN and fuzzy logic. Some of them are reported in [3-11]. Such structures usually outperform LTE and also compensate for nonlinearities in the channel with varying degree of success. ANN based equalizers work even if the channel is unknown while fuzzy equalizers require that the channel must be known a prior. The major problem of such equalizers is the high complexity and computational requirement which can further be reduced. An integration of ANNs and fuzzy set theoretic approach will offer advantages of both the techniques.

The common neuron model has been modified to obtain a generalized neuron (GN) model using fuzzy compensatory operators to reduce the complexity of the structure and overcome the problems such as initial selection of architecture of neural network giving optimum performance for complex function mapping, which affects the training time requirement and also fault tolerant capabilities of the ANN [12]. This neuron provides flexibility and fault tolerant capability to cope up with the nonlinearities involved in the system. GN has been used successfully for power systems problems [13-14]. Application of this significantly reduced complexity GN as a channel equalizer in digital communication systems is demonstrated in this paper.

There have been introduced many different nonlinear devices models and channel models, so a unitary comparison between all known equalizers is difficult to be done. Bayesian solution provides the minimum average BER achievable for symbol decision and indirect-modeling equalizer structure. It has been shown that BER of proposed equalizer outperforms conventional LTE LMS equalizer approaches to that of optimal Bayesian equalizer under linear and non-linear conditions. The proposed equalizer provides acceptable training and BER performance.

2. GENERALIZED NEURON MODEL

Existing conventional neuron model generally uses an aggregation function and its transformation through an activation function. Generally summation is used as aggregation and sigmoid, radial basis, tangent hyperbolic or linear limiters etc. as activation function. Generalized neuron structure shown in figure 3 is developed by modifying the conventional neuron structure using fuzzy compensatory aggregation operators along with fuzzy activation functions. Aggregation operation in GN is performed partly by sum ($\Sigma$) and partly by product ($\prod$) functions. Bipolar sigmoid function ($f_1$) is used as a transformation function for $\Sigma$ part and Gaussian function ($f_2$) is used as transformation function for $\prod$ part of the structure. The final output is the summation of the $\Sigma$ output and $\prod$ output with weight sharing as $W$ and $(1-W)$ respectively. The input output relationship for GN is given by following expressions.

Let $Y$ represents the input vector to the equalizer which is $y[k]$ as given by equation (2).

The $\Sigma$ part output of GN is calculated as

$$Os = 2s_{out} - 1$$

(8)
Where, \( S_{out} = \frac{1}{1 + e^{-\lambda s \cdot \text{sum}_s}} \).  

Figure 3: Structure of GN

\( \lambda_s \) is the gain scale factor for \( \Sigma_1 \) part of GN.

\( X_{os} \) is the bias to \( \Sigma_1 \) part and \( W_{si} \) is the weight vector.

The \( D \) part output of GN is calculated as

\[ Op = e^{-\lambda_p \cdot \text{prod} - p^2} \]  

Where, \( \lambda_p \) is the gain scale factor for \( D \) part of the GN.

\( X_{op} \) is the bias to \( D \) part and \( W_{pi} \) is the weight vector.

The final output is

\[ O_i = W \cdot O_s + (1-W) \cdot Op \]  

Where, \( W \) is the weight vector. \( O_i \) is the estimated output vector \( s'[k-d] \).

3. Learning Algorithm of GN

Back propagation learning algorithm is used to train the network.

Following steps are to be followed to train the network till the mean square error reaches to minimum.

Step 1 Calculate the output for each pair of input using equation no. 8, 10 & 11.

Step 2 Calculate the error using the following relation

\[ E_i = (O_i - D_i) \]  

Where, \( D_i \) is the desired output \( s[k-d] \).

Step 3 Calculate the mean square error for convergence as

\[ \frac{N}{5.0} \sum E_i^2 \]  

Where, \( N \) is the total number of training patterns and a multiplication factor of 0.5 has been taken to simplify the calculations.

Step 4 Different weights of the networks are updated as following

(a) Weights associated with \( \Sigma_1 \) and \( \Sigma_2 \) part of the GN are updated as \( W(k) = W(k-1) + \Delta W \)

Where, 
\[ \Delta W = \eta \delta_i (O_s - Op) + \alpha W(k-1) \]  

and 
\[ \delta_i = \sum (O_i - D_i) \]

(b) Weights associated with inputs and \( \Sigma_1 \) part of the GN are updated as \( W_{si}(k) = W_{si}(k-1) + \Delta W_{si} \)

Where, 
\[ \Delta W_{si} = \eta \delta_i Y_i + \alpha W_{si}(k-1) \]

And,

(c) Weights associated with inputs and \( D \) part of GN are updated as \( W_{pi}(k) = W_{pi}(k-1) + \Delta W_{pi} \)

Where, 
\[ \Delta W_{pi} = \eta \delta_p Y_i + \alpha W_{pi}(k-1) \]  

and 
\[ \delta_p = \sum \delta_i (1-W) \ast (-2 \ast \text{prod} - p) \ast Op \]  

Where, \( \eta \) is the learning rate and \( \alpha \) is the momentum factor, whose values ranges between 0 and 1 determined by trial and error.
4. **Simulation Results and Discussion**

Following channel model is used to simulate the channel.

\[ H(z) = 0.3482 + 0.8704 z^{-1} + 0.3482 z^{-2} \]  

(17)

For linear channel \( NL=0 \), \( y'[k] = x[k] \) \hspace{1cm} (18)

The nonlinear channel, \( NL=1 \) is modeled to nonlinearity introduced due to saturation of amplifiers used in the transmission systems as

\[ y'[k] = \tanh(x[k]) + q[k] \]  

(19)

\( NL = 2 \) is modeled to random nonlinear distortions as

\[ y'[k] = x[k] + 0.2 x^2[k] - 0.1 x^3[k] + q[k] \]  

(20)

Fourth order, \( m=4 \) equalizer with delay, \( d=1 \) is simulated.

Channel equalizer is implemented using GN. For training a random sequence of 1000 duobinary signals of \{1, -1\}, equi-probable and independent identically distributed (i.i.d) is generated and passed though the channel. Nonlinearities and white Gaussian noise are further introduced. Initial weights are generated randomly. This generated sequence is used to train the equalizer with back propagation learning algorithm for 300 epochs to obtain minimum mean square error. Values of \( \eta \) and \( \alpha \) are chosen as 0.0015 and 0.5 respectively.

Fig (4) shows the convergence characteristics for the three nonlinear channel models for SNR=16 dB. The characteristics show fast and smooth convergence of error for all the three nonlinear channel models and make it suitable for real time applications.

The trained equalizer is tested using separately generated duo-binary, equi-probable and i.i.d. sequence with nonlinearities and white Gaussian noise added. The results are averaged over 10 repetitions using testing sample of size 10000 each. SNR is varied between 2-20 dB in steps of 2 dB to ascertain performance under different noise conditions. Fig (5-7) shows the plot of BER performance of channel for the channel with \( NL=0 \), \( NL=1 \) and \( NL=2 \) respectively which clearly shows the capability of the GN based equalizer to reconstruct the received destroyed signals. The BER performance is also compared with conventional LTE LMS equalizer and optimal Bayesian solution. BER of the proposed equalizer outperforms conventional filter based LMS equalizer and approaches to Bayesian performance even when the equalizer is operated under severe nonlinearity and low SNR conditions. Superior performance of GN based equalizer over linear LMS equalizer for all the three channel models is quite evident from the figures 5-7. There is severe BER performance degradation in the conventional LMS equalizer as the severity of nonlinearity increases while the BER performance of the GN based equalizer is quite similar to each other for both...
the linear and nonlinear channel models and approaches to optimal Bayesian solution especially for severe nonlinearity NL=2 throughout wide variation of SNR.

5. CONCLUSION
Computationally efficient GN based equalizer is described. There is no problem of selection of initial architecture of neural network as only a single GN is required. No hidden layer is required. Reduced complexity and much simple design procedure is required. Fast convergence of error during training is achieved because it has a much smaller number of weights to be adjusted hence suitable for real time applications. The simulation results show that this neuron provides flexibility and fault tolerant capability to cope up with the nonlinearities involved and that proposed equalizer BER performance outperforms conventional filter based LMS equalizers and approaches to optimal approaches to optimal Bayesian solution for both linear and nonlinear conditions. GN based equalizers offer the advantages of both reduced complexity and good acceptable BER performances hence attractive alternatives for designing equalizers for digital communication systems.

REFERENCES


Authors’s Biography

Vineeta Choudhary is currently working as a Reader in Electronics and Communication Engineering, at Ujjain Engineering College, Ujjain, M.P. Her area of interest is digital communications, soft computing.

Dr. B. K. Joshi is currently working as Professor in Electronics and Telecommunication Engineering and Computer Engineering, at Military College of Telecommunication Engineering, Mhow, M.P. His area of interest is programming languages, compiler design, digital communications, mobile adhoc and wireless sensor networks and software engineering.
Traffic Analysis of Message Flow In Three Crossbar Architecture Space-Division Switches

D. Shukla\textsuperscript{1}          R. Singhai\textsuperscript{2}

\textbf{Abstract}

In computer networks the space division switches are used to transmit the messages in order to reach its destinations. These switches are based on cross bar technology. Actually, it is built up of several smaller rectangular crossbars and less cross points is needed than the traditional cross bar switches. According to Tananbum (1996) if we increase the number of cross points, the outgoing reaching probability of messages also increases accordingly with the high cost and low congestion in the network. In this paper we considered the architecture of a three crossbar space-division switch and with the help of Markov chain theory, an L-dependent mathematical model is proposed and used to calculate reaching probabilities of message flow.

\textbf{Keywords} : Space-Division Switch, Cross-Bar Technology, Markov Chain Model, Reaching Probabilities, Transition Probability Matrix, Simulation Study, Message Flow.

1. \textbf{Review Of Literature}

2. **Motivation**

Shukla and Gadewar [11] have suggested a Markov chain model for the transitional analysis of message flow in a two crossbar space division switches. We extend this model, in this paper, from two-crossbars to three crossbar setup and with the help of a simulation study, the impact on reaching probabilities of message is analyzed.

3. **Introduction and Assumptions**

In what follows, we consider a space-division switch [11],[12] with parameters $N = 16, n = 4, t = 3$ shown in fig. 1 and assume the followings:

a) The left side of switches is input and the flow of information is from left to right.

b) Each input line, on left side, is attached with a computer having different initial probabilities of selection by users. This level is the stage 1.

c) The middle crossbars are stage 2 containing three crossbars with each having four inputs and four output lines.

d) The third stage contains four crossbars, each with three inputs and four output lines. At this, three output lines are with computers and the fourth one, in each crossbar, is a loss state.

e) The term $I(M,K,L)$ denotes an input state at $M$th stage in $K$th crossbar and at $L$th input line where $M=1,2,3; K=1,2,3,4; L=1,2,3,4$. For example, in fig. 1 the term $a_{1}$ is input state $I(1,1,1), a_{2}$ is state $I(1,1,2), c_{1}$ is $I(1,2,1), e_{1}$ is $I(2,1,1), g_{1}$ is $I(2,2,1)$, and $i_{1}$ is $I(3,1,1)$.

f) The term $O(M,K,L)$ denotes output state at $M$th stage, in $K$th crossbar and $L$th output line like the term $b_{1}$ is output state $O(1,1,1), b_{2}$ is $O(1,1,2), b_{3}$ is $O(1,1,3), d_{1}$ is $O(1,2,1), f_{1}$ is $O(2,1,1), h_{1}$ is $O(2,2,1)$ and $j_{1}$ is $O(3,1,1)$.
As special, the output states O(3,1,4), O(3,2,4), O(3,3,4) and O(3,4,4) are loss states and when a message reaches to them, it is assumed lost or reached to the known destinations.

3.1. Markov Chain Model

Let \( \{X_n : n = 0, 1, 2, 3, \ldots \} \) be a Markov chain with state space \( I(M,K,L) \) and \( O(M,K,L) \), \( M=1,2,3 \) and \( K,L = 1,2,3,4 \). The \( X_n \) denotes the state of message at the \( n \)th step transition over states \( I(M,K,L) \) and \( O(M,K,L) \). The unit-step transition probabilities over states are:

\[
P[X_{n+1} = I(1, K, L) | X_n = I(1, K, L)] = \text{L}_{1k}
\]

\[
P[X_{n+1} = I(1, K, L) | X_n = O(1, K, L)] = 1 - \sum_{i} \text{P}_{ik}
\]

\[
P[X_{n+1} = I(2, K, L) | X_n = O(2, K, L)] = \text{P}_{1k}
\]

When \( L=1,2,3; K=1,2,3,4 \)

\[
\text{P}_{ik} = \frac{1}{(M-1)!} \begin{vmatrix}
0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 0 \\
0 & 0 & 0 & 0 & \ldots & 0
\end{vmatrix}
\]

The terms \( \text{L}_{ik}, \text{P}_{ik}, \text{Q}_{ik}, \text{R}_{ik} \) \( (i=1,2,3) \) are the probabilities of transition lying between 0 and 1 and placed as elements of transition probability matrix given on the next page.

Table I: Transition Probability Matrix (t. p. m.) for Stage 1

<table>
<thead>
<tr>
<th>States</th>
<th>( I(0, K, 0) )</th>
<th>( I(0, K, 2) )</th>
<th>( I(0, K, 3) )</th>
<th>( I(1, K, 0) )</th>
<th>( I(1, K, 2) )</th>
<th>( I(1, K, 3) )</th>
<th>( O(0, K, 1) )</th>
<th>( O(0, K, 2) )</th>
<th>( O(0, K, 3) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I(0, K, 0) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\text{L}_{ik}</td>
<td>\text{L}_{ik}</td>
<td>\frac{1}{L} \text{(L}<em>{ik} + \text{L}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I(0, K, 2) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\text{L}_{ik}</td>
<td>\text{L}_{ik}</td>
<td>\frac{1}{L} \text{(L}<em>{ik} + \text{L}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I(0, K, 3) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\text{L}_{ik}</td>
<td>\text{L}_{ik}</td>
<td>\frac{1}{L} \text{(L}<em>{ik} + \text{L}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I(1, K, 0) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\text{L}_{ik}</td>
<td>\text{L}_{ik}</td>
<td>\frac{1}{L} \text{(L}<em>{ik} + \text{L}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I(1, K, 1) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\text{L}_{ik}</td>
<td>\text{L}_{ik}</td>
<td>\frac{1}{L} \text{(L}<em>{ik} + \text{L}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I(1, K, 2) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\text{L}_{ik}</td>
<td>\text{L}_{ik}</td>
<td>\frac{1}{L} \text{(L}<em>{ik} + \text{L}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I(1, K, 3) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\text{L}_{ik}</td>
<td>\text{L}_{ik}</td>
<td>\frac{1}{L} \text{(L}<em>{ik} + \text{L}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( O(0, K, 0) )</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\frac{1}{P} \text{(P}<em>{ik} + \text{P}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( O(0, K, 1) )</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\frac{1}{P} \text{(P}<em>{ik} + \text{P}</em>{ik})</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( O(0, K, 2) )</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\text{P}_{ik}</td>
<td>\frac{1}{P} \text{(P}<em>{ik} + \text{P}</em>{ik})</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1562
3.2 Model Classification

The probabilities \( L_{ik} \), \( P_{ik} \), \( Q_{ik} \), \( R_{ik} \) and \( S_{ik} \) may be functions of \( M \), \( K \) and \( L \) parameters and on this basis the classification of Markov chain models be as below:

(i) **M-Dependent model** - where probabilities \( L_{ik} \), \( P_{ik} \), \( Q_{ik} \), \( R_{ik} \) and \( S_{ik} \) are only functions of \( M \).

(ii) **K-Dependent model** - where probabilities \( L_{ik} \), \( P_{ik} \), \( Q_{ik} \), \( R_{ik} \) and \( S_{ik} \) are only functions of \( K \).

(iii) **L-Dependent model** - where probabilities are functions of \( K \) and \( L \) parameters both.

4. **Calculation Of Reaching (Initial) Probabilities**

Let \( P_{ik} \) \( (i = 1, 2, 3) \) be the probability of choosing the \( i^{th} \) input line in \( K^{th} \) switching element of the space division switch configuration given in fig. 1of the section 1.0.

For \( i = 4 \), the probability is \( \sum_{k=1}^{3} p_{ik} \). For the Markov chain \( \{X_n, n = 0, 1, 2, 3, \ldots\} \) over the states \( I(M, K, L) \), the initial probabilities of choosing a connecting path is

\[
P(X_0 = I(1, K, 1)) = p_{11} \quad \text{and} \quad P(X_0 = I(1, K, 3)) = p_{13}.
\]
4.1. Outgoing Probabilities at Stage 1, 2 and 3

The O(1,K,L) over varying K and L are the outgoing states, for the stage 1, where the message is ready to route into for the next stage.

\[ P[ X_1 = O(1,K,L) ] = P[ \text{message reaches to the state } O(1,K,L) \text{ at the first step} ] \]

The general form for \( M = 1 \) (stage-1) is

\[ P[ x = O (1, k, L) ] = L_1^k \text{ when } L = 1; K = 1, 2, 3, 4 \]
\[ = L_1'^k \text{ when } L = 2 \]
\[ = \{1 - L_1^k + L_2^k\} \text{ when } L = 3 \]

The general form for \( M = 2 \) (stage-2) is

\[ P[ x = O (2, k, L) ] = Q L_k \sum_{i=1}^{L} L_{ii} \text{ when } k = 1, L = 1, 2, 3 \]
\[ = \{1 - L_1 k + L_2 k\} \text{ when } k = 1, L = 4 \]
\[ = Q L_k \sum_{i=1}^{L} L_{ii} \sum_{i=1}^{L_{ii}} \text{ when } k = 2, L = 1, 2, 3 \]
\[ = \{1 - L_1 k + L_2 k\} \{4 - L_{ii} - \sum_{i=1}^{L_{ii}} \text{ when } k = 2, L = 4 \]
\[ = R L_k \sum_{i=1}^{L} L_{ii} \sum_{i=1}^{L_{ii}} \sum_{i=1}^{L_{ii}} \text{ when } k = 3, L = 1, 2, 3 \]
\[ = \{1 - L_1 k + L_2 k\} \{4 - L_{ii} - \sum_{i=1}^{L_{ii}} \text{ when } k = 3, L = 4 \]

The general form for \( M = 3 \) (stage-3) is

\[ P[ x = O (3, k, L) ] = R L_k \]
\[ = \{1 - L_1 k + L_2 k\} \{4 - L_{ii} - \sum_{i=1}^{L_{ii}} \text{ when } k = 4, L = 1, 2, 3 \]

5. L-Dependent Model and Simulation Study

Based on the above equations we considered the following \( L \)-dependent Markov chain model with unit-step transition probability. The \( a, b, c, d \) and \( e \) are constants having values in between 0.00 to 0.5.
Traffic Analysis Of Message Flow In Three Crossbar Architecture Space-division Switches

5.1 Effect of L And D

The fig. 5.1 to 5.3 shows the variations over the reaching probabilities with respect to increasing values of a, b, c, d parameters.

When L=1 and a=c=d=0.1, the reaching probability P[M,K,1] reduces for the increasing values of K. This connectivity probability is higher for stage 1 and suddenly decreases for other stages. When L=2, the connectivity probability shows different pattern of variations than to compare what shown at L=1. The probability P[M,K,2] is higher for stage 1 and shows sudden decreasing pattern for other stages. When we talk about L=3, the P[M,K,3] has low down tendency over increasing K. The probability of connectivity improves at the third stage in comparison to second.

5.2 Effect of Variation of C, D and K

Figure 5.1 : a=0.1,c=0.1,d=0.1, L=1

Figure 5.2 : a=0.1,c=0.1,d=0.1, L=2

Figure 5.3 : a=0.1,c=0.1,d=0.1, L=3

Figure 5.4 : a=0.1,c=0.5,d=0.1, L=1

Figure 5.5 : a=0.1,c=0.5,d=0.1, L=2

Figure 5.6 : a=0.1,c=0.5,d=0.1, L=3

Fig. 5.4 to 5.6 shows the effect on the reaching probabilities with the increasing values of c and d parameters while keeping the value of a parameter constant. With the increase of C, the probability pattern

\[ p[X_i=O(K,J)X_i=O(K,L)] = \begin{cases} 1-2(a)^{2} & J=1,2,3,4 \\ 1-6(b)^{2} & J=4 \\ 1-6(c)^{2} & J=1,2,3,4 \\ 1-6(d)^{2} & J=4 \\ 1-2(e)^{2} & J=1,2,3,4 \\ 1-6(f)^{2} & J=4 \\ 1-6(g)^{2} & J=1,2,3,4 \\ 1-6(h)^{2} & J=4 \end{cases} \]

Figure 5.4: a=0.1,c=0.5,d=0.1, L=1

Figure 5.5: a=0.1,c=0.5,d=0.1, L=2

Figure 5.6: a=0.1,c=0.5,d=0.1, L=3
at stage 2 bears a sudden change. When \( L=1 \), the first and third stage has downward trend of probabilities over varying \( k \), but the third stage bears a little increase than the second stage. When \( L=2 \), the first stage remains at high probability in comparison to others. At \( L=3 \) the chance of reaching probabilities reduces for the third stage with increasing values of \( K \), but this increases for the second stage over the same \( K \). The increase in \( c \) has special impact on the second stage probability of connectivity. When \( C \) is high, the probability for \( M=2 \) concentrate entirely near to \( K=1 \).

5.3 Effect Of Variation Of \( A,C,D \) and \( L \)

With the increase in parameter \( a \), in comparison to \( c, d \) and \( L \) has an effect in the connectivity probability. According to figure 5.7 to fig. 5.10, the increase of \( k \) produces decreasing Probability \( P[M,K,L] \). When \( L=1 \) changes to \( L=2 \), keeping fix \( a, c, d \), we observe higher probability \( P[M,K,2] \) than \( P[M,K,1] \). When \( L=3 \), the \( K=2 \) is an ultimate value.

With the increase in \( d \) value, from \( d=0.1 \) to \( d=0.5 \), a sudden increase of \( P[M,K,L] \) is observed at \( k=3, k=4 \). In this, \( L=3 \) bears the largest probability as showing in fig. 5.11 to 5.12. The increase in \( c \) Value produces zigzag movement in \( P[M,K,L] \). The \( L=1 \) and \( L=3 \) has a sudden jump in connectivity probability. With the simultaneous increase in \( c \) and \( d \) value, the connectivity with \( M=3 \) increases at high rate. In this case, the \( K=3 \) bears the highest probability for \( P[M,K,3] \). The increase in value of a parameter has the most significant, effect in the role of increasing the connectivity probability. According to fig 5.13, when all the parameters \( a, c, d, L \) are high than

\begin{figure}
\includegraphics[width=0.5\textwidth]{fig5.7.png}
\caption{Figure 5.7: \( a=0.2, c=0.1, d=0.1, L=2 \)}
\end{figure}

\begin{figure}
\includegraphics[width=0.5\textwidth]{fig5.8.png}
\caption{Figure 5.8: \( a=0.2, c=0.1, d=0.1, L=1 \)}
\end{figure}

\begin{figure}
\includegraphics[width=0.5\textwidth]{fig5.9.png}
\caption{Figure 5.9: \( a=0.2, c=0.1, d=0.1, L=2 \)}
\end{figure}

\begin{figure}
\includegraphics[width=0.5\textwidth]{fig5.10.png}
\caption{Figure 5.10: \( a=0.2, c=0.1, d=0.1, L=3 \)}
\end{figure}

\begin{figure}
\includegraphics[width=0.5\textwidth]{fig5.11.png}
\caption{Figure 5.11: \( a=0.2, c=0.5, d=0.1, L=1 \)}
\end{figure}

\begin{figure}
\includegraphics[width=0.5\textwidth]{fig5.12.png}
\caption{Figure 5.12: \( a=0.2, c=0.5, d=0.1, L=3 \)}
\end{figure}

With the increase in \( d \) value, from \( d=0.1 \) to \( d=0.5 \), a sudden increase of \( P[M,K,L] \) is observed at \( k=3, k=4 \). In this, \( L=3 \) bears the largest probability as showing in fig. 5.11 to 5.12. The increase in \( c \) Value produces zigzag movement in \( P[M,K,L] \). The \( L=1 \) and \( L=3 \) has a sudden jump in connectivity probability. With the simultaneous increase in \( c \) and \( d \) value, the connectivity with \( M=3 \) increases at high rate. In this case, the \( K=3 \) bears the highest probability for \( P[M,K,3] \). The increase in value of a parameter has the most significant, effect in the role of increasing the connectivity probability. According to fig 5.13, when all the parameters \( a, c, d, L \) are high than
the connectivity chances are also high, for both $M=2$ and Therefore, the higher values of parameter $a$, $c$, $d$ produces higher chance of connectivity and message passing in space division switches having three cross bar.

6. Conclusions

Many interesting highlights are identified after the simulation study on different values of parameters using L Dependent Model. Some concluding remarks are:

(i) When parameters are with smaller values, the outgoing reaching probability for $L=2$ is higher than $L=1$ at $K=1$, i.e. $(P(M,K,2)>P(M,K,1))$. The increase in $d$ values (from 0.1 to 0.5) certainly affects the outgoing probabilities in $L$-dependent model. So one should be optimistic and careful in selecting $d$ values. However, the variation of parameter $c$ and $d$ both affects the reaching probabilities and their high values (e.g., $c=0.3$, $d=0.3$) produce a significant change in them. In three cross bars, there is constant exponential decay in $P(M,K,L)$ is found. With the increase of $d$ value this decay process reduces, therefore the higher value of $d$ is recommended in case of three pin cross bars.

(ii) In $L$-dependent model, the increase in parameter $a$ plays very important role and has significant impact on outgoing probability.

(iii) The parameter has very important role in deciding about the probability pattern of outgoing message. With the increase of $c$ and $d$ values together produces significant increase in message passing probability in case of three crossbar setup. When all values of $a$, $c$ and $d$ are high i.e. in the range (0.3, 0.6), the $L$-dependent model shows better performance in this case than two-cross bar setup.

(iv) One interesting observation in three crossbar case is, that for $L=4$. The reaching probability is much higher. So, with the help of proposed model, the hardware designers of space division switches can design switches, more effectively & efficiently.

References


[5] Li.S.Q, “Nonuniform traffic analysis on a nonblocking space-division packet switch”, IEEE


Author’s Biography

Dr. Diwakar Shukla is presently working as a faculty member in the Department of Mathematics and Statistics, H.S.Gour Sagar University, Sagar, M.P and having over 19 years experience of teaching to U.G and P.G classes. He obtained M.Sc.(stat.), Ph.D.(stat.) degrees from Banaras Hindu University, Varanasi and served the Devi Ahilya University, Indore, M.P as a permanent Lecturer from 1989 for nine years and obtained the degree of M.Tech.(Computer Science) from there. He joined Sagar University,Sagar as a Reader in statistics in the year 1998. During Ph.D. from BHU, he was junior and senior research fellow of CSIR, New Delhi through Fellowship Examination (NET) of 1983. Till now, he has published more than 55 research papers in national and international journals and participated in
more than 35 seminars/conferences at the national level. He is the recipient of MPCOST Young Scientist Award, ISAS Young Scientist Medal, UGC Career Award and UGC visiting fellow to Amerawati University, Maharashtra. He also worked as a Professor in the Lucknow University, Lucknow, U.P., for one (from June, 2007 to 2008) year and visited abroad to Sydney (Australia) and Shanghai (China) for conference participation and paper presentation. He has supervised nine Ph.D. theses in Statistics and Computer Science and six students are presently enrolled for their doctoral degree under his supervision. He is a member of 10 learned bodies of Statistics and Computer Science at the national level. The area of research he works are Sampling Theory, Graph Theory, Stochastic Modeling, Data mining, Operation Research, Computer Network and Operating Systems.

Rahul Singhai has obtained M.C.A. degree from H.S. Gour University, Sagar, MP, in 2001 and obtained M.Phil degree in Computer Science from Madurai Kamaraj University, Madurai, Tamilnadu in 2008. Presently he is pursuing Ph.D in Computer Science from H.S. Gour University, Sagar. His research interest includes Computer Network, Data mining & Software Testing. He has authored and co-authored 8 research papers in proceedings & journals. Currently, he is working on to develop new probability based methods for data preprocessing in data mining. He has worked as contract Lecturer in the Dept. of Computer Science & Applications., H.S. Gour University, Sagar from Feb -2005 to July-2009. He is presently working as permanent lecturer at IIPS, Devi Ahilya Vishwavidyalaya, Indore (M.P.) since 7th July-2009.
Modelling and Dynamic Analysis of Pneumatic Control Valve with Stiction

S.Sivagamasundari¹ D.Sivakumar²

Abstract
High performance of control loops is necessary to ensure high product quality and low cost in chemical plants. Often poor performance can be detected by the operator, but many times the problem may propagate to other loops, making it difficult to detect the root cause. One of the many reasons for degrading performance is stiction found in control valves. It hinders the proper movement of valve stem and consequently affects the control loop performance. It is therefore important for control engineers to understand stiction phenomena and to know how to deal with them. This paper focuses on the development of a model for control valve stiction, through simulation, which can be used for the dynamical analysis of valve stiction effects on process control loop performance. The model developed here use simple empirical relationships between controller output and the valve position to describe valve stiction with just a few parameters that can be determined from operating data. The proposed stiction model captures the stiction behaviour and can be used to investigate by analysis and simulation, the properties of stiction that are relevant to control design.

Keywords: Control Valve, Stiction, Limit Cycle, Dead Band, Hysteresis, Slip Jump

1. Introduction
The field of controller performance monitoring has received much attention in the engineering research literature. However the diagnosis of poor performance remains an open area. The reason for poor control loop performance may be due to poor controller tuning, presence of disturbances, process and/or actuator non linearities. Non linearities degrade the performance of the controller in several ways. They may produce oscillations in process variable, shorten the life of control valve, may upset process stability and in most cases lead to inferior quality end products, thus causing larger reduction rates and reduced profitability.

The non-linearities may be present in the process itself or in the actuators or control valves. The most popular problem is the incorrect tuning of controllers which may be due to non-linearities present in the process itself, can be faced by process identification and auto tuning. Instead, a more common cause should be sought in the presence of static friction in actuators, which causes a delayed and sluggish actuation of changes in manipulated variables, required by the control system. Stiction found in pneumatic control valves is one of the greatest obstacles in high precision process control systems. It can cause steady state and tracking errors, while it may result in limit cycles. Therefore its influence on the response of systems must be seriously considered.

¹Department of Instrumentation Engg., Annamalai University, Annamalainagar, India. Email : sivagasundari67@gmail.com
²Department of Instrumentation Engg., Annamalai University, Annamalainagar, India. Email : dsk2k5@gmail.com
Control strategies that attempt to compensate for effects of stiction inherently require a suitable model to predict and to compensate for the stiction. A good stiction model is also necessary to analyse stability, predict limit cycles, find controller gains and perform simulations. Both detailed physical models and purely empirical models have been used to simulate valve stiction. Physical models [1] describe the stiction phenomenon using force balance based on Newton’s second law of motion. The main disadvantage of these models is that they require knowledge of several parameters such as the mass of moving parts and different types of friction forces that cannot be easily measured and are dependent on the type of fluid and valve wear. On the other hand empirical or data driven models [2] use simple empirical relationships between controller output and the valve position to describe valve stiction, with just a few parameters that can be determined from operating data. Srinivasan et al.[3] uses a Hammerstein model identification approach along with one parameter stiction model [4] to detect and quantify stiction. The one parameter model does not catch the true stiction behaviour. Choudhury et al.[1,2] have discussed the definition of stiction thoroughly, distinguished it from other valve nonlinearities, and proposed a new two parameter data driven model of stiction. The model derived by Choudhury [2] has been widely used in the study of valve stiction. In this work the complexity of Choudhury’s model is reduced and another data driven model is proposed with straightforward logic flow.

The paper is organized as follows. Section 2 recalls the basic structure of a control valve system. Stiction behaviour is explained in section 3. The physical model of stiction is given in section 4. In section 5, the new data driven valve stiction model is proposed. Based on this model the effects of control valve stiction on a control loop are illustrated. The application of the proposed model to simulated example is presented in section 6. Section 7 analyses the effects of stiction under closed loop condition. Conclusion is given in section 8.

2. THE STRUCTURE OF A CONTROL VALVE

The general structure of a pneumatic control valve is shown in fig.1. The valve is closed by elastic force and opened by air pressure. Flow rate is changed according to the plug position which is determined by the balance between elastic force and air pressure.

The plug is connected to the valve stem. The stem is moved against static or kinetic frictional force caused by packing, which is a sealing device to prevent leakage of process fluid. Smooth movement of the stem is restricted by excessive static friction. The valve position cannot be changed until the controller output overcomes static friction, and it is suddenly and considerably changed when the difference between elastic force and air pressure exceeds the maximum static frictional force.

3. STICTION BEHAVIOUR

Stiction (also known as stick-slip or static friction) in control valves is thought to occur due to seal degradation, lubricant depletion, inclusion of foreign matter, activation of metal sliding surfaces at high temperatures and tight
packing around the stem. The resistance offered from the stem packing is often cited as the main cause of stiction. One other very common cause of stiction is indirectly due to regulations on Volatile Organic Compound (VOC) emissions. In many plants, a team monitors each valve for VOC emissions, usually between the packing and the stem. If any minute leakage is detected, packing in the valve body is tightened, but tightened far more than is necessary. This causes the valve to stick, making the process run less efficiently with increased energy consumption. Stiction often varies over time and operating regimes. Since wear is also non-uniform along the body, frictional forces are different at different stem positions, when the control loop is at steady state, and if a valve exhibits this behaviour, persistent oscillations in process variable on either side of the set point are observed.

Understanding the type of oscillations caused by a sticking valve in a control loop requires a good grasp of the stick-slip phenomenon. Stiction in control valves leave a distinct qualitative shape in the controller output and process variable. These shapes can be generally categorized as being square, triangular or saw toothed and depends primarily on the type of controller structure implemented. Under some condition, a control valve will exhibit stick-slip behaviour as shown in fig. 2.

The input force (control action) observed during motion is sketched. During the stick, (interval a-b) the force rises. At point ‘b’, the force reaches \( F_s \), the level of static friction when the system has been at rest for considerable time, and slip begins. During interval b-c slip occurs. At point ‘c’, the pin is arrested and the spring force again begins to rise entering a stable limit cycle c-d-e. Here \( F_s \) is the static friction and \( F_c \) is the coulomb friction.

4. PHYSICAL MODEL OF VALVE STICTION

For a pneumatic sliding stem valve, the force balance equation based on Newton’s second law can be written as.

\[
M \frac{d^2x}{dt^2} = \sum \text{Forces} = F_s + F_c + F_p + F_i
\]  

(1)

Where,

- \( M \) = Mass of the moving parts;
- \( x \) = Relative stem position;
- \( F_s \) = \( S_u \) = Force applied by pneumatic actuator
- \( S_u \) = Area of the diaphragm,
- \( u \) = Valve input signal (controller output))
- \( F_c = -Kx \) = spring force (K = Spring constant)
- \( F_p = -\alpha \Delta p \) = force due to fluid pressure drop \( \alpha \) = plug unbalance area,
- \( \Delta p \) = fluid pressure drop across the valve
- \( F_i \) = Extra force required to force the valve to be into the seat
- \( F \) = Frictional force.

Here \( F_s \) and \( F_p \) assume to be zero because of their negligible contribution in the model. Now the force balance equation can be written as

\[
M \frac{d^2x}{dt^2} = \sum \text{Forces} = \dot{x} - kx - F_i
\]  

(2)

\[
M \frac{d^2x}{dt^2} = \sum \text{Forces} = \dot{x} - kx - F
\]  

(3)

Where, \( x \) = Position of the stem;
\[ x_2 = v = \text{Stem velocity} \quad (4) \]

\[ F = \begin{cases} F(v) & \text{if } v \neq 0 \\ F_s & \text{if } v = 0 \quad \& \quad |F_s| < F_c \\ F_c \, \text{sgn}(F_c) & \text{if } v = 0 \quad \& \quad |F_c| \geq F_c \\ \end{cases} \quad (5) \]

Where,

\[ F(v) = F_c \left( \text{sgn}(v) \right) + \frac{F_s}{|F_c|} \left( v \right) \]

The equations (1-6) describe the valve model with friction forces.

Here, \( F_v = \text{Viscous friction coefficient}, \) \( F_s = \text{Static friction}, \) \( F_c = \text{Coulomb friction}, \) \( v_s = \text{Stribeck’s constant}. \)

The function ‘\( F \)’ is easily obtained by measuring the friction force for motions with constant velocity.

A disadvantage of a physical model of a control valve is that it requires several parameters to be known. The mass, ‘\( m \)’, spring constant ‘\( K \)’ and typical friction forces depend upon the design of the valve.

5. The Data Driven Model Formulation

Fig.3 shows the typical input output behaviour of a sticky valve. Without stiction the valve will behave like a linear system. (any amount of valve input would result in the same amount of valve output). After the controller output overcomes the dead band (AB) and stick band (BC) of the valve, the valve jumps to a new position (D) and continues to move (DE). Due to very low or zero velocity, the valve may stick again in between points E and F while travelling in the same direction. In such a case, the magnitude of dead band is zero and only stick band(S) is present. This can be overcome if the controller output signal is larger than stick band only. The dead band and the stick band represent the behaviour of the valve when it is not moving, though the input to the valve keeps changing. Slip jump (J) represents the abrupt release of potential energy stored in the actuator chamber due to high static friction. The magnitude of slip jump is very crucial in determining the limit cyclic behavior introduced by stiction.

However for a sticky valve static and dynamic frictions must be taken into account. The dynamic friction band \( f_D \) is given by \((S-J)/2 \) and static friction band \( f_s \) is given by \((S+J)/2 \); where J is the slip jump and S is the sum of stick band and dead band. The slip jump is equal to stick band. Based on the sticky valve behaviour a new valve stiction model is proposed. The valve sticks only when it is at rest or it is changing its direction it comes to a rest momentarily. Once the valve overcomes stiction, it starts moving and may keep on moving for sometime depending on how much stiction is presenting the valve. In this moving phase, it suffers only dynamic friction which may be smaller than the static friction. it continues to do so until its velocity is again very close to zero or it changes direction.

Figure 3: Typical Stiction Behaviour
Figure 4: Flow Chart Of The Proposed Valve Stiction Model

Figure 4 shows the flow chart of the proposed valve stiction model. Here u(t) is the controller output,

\[ u(t) = (u(t) - u(t-1)) \]

Yes \[ |u(t)| > f_s \] No

\[ x(t) = u(t) - (\text{sign}(u(t)) - f_d(t)) \]

6. VALVE SIMULATION

The purpose of simulation is to determine the influence of the friction terms in the model. The non linearity in the model is able to induce limit cycles in the feedback control loop. The valve model is driven by a periodic ramp signal in open loop. During closed loop condition, input to the valve is given from controller, so controller output is acting as input to the valve stiction model and valve position is the output of the model.

A. Case(i): Linear

Using the aforesaid model, the response of the control valve to a periodic ramp signal in the absence of stiction \( (f_s = f_d = 0.0001 \text{ (closer to zero))} \) is obtained and is shown in fig.5. This can be assumed as a linear valve in the absence stiction. The step response of the valve in the absence of stiction is shown in Fig.6. The response is similar to that of a first order system with no offset.

B. Case(ii): Pure Dead band

When the stick band is zero, there is no slip jump and \( f_s = f_d \). In this condition only dead band arises because on changing the direction, the valve remains stationary until the net applied force is large enough to overcome the static friction. If the static friction is larger dead band will also be larger. This condition is shown in Fig.7. The step response for this condition will be similar to case (i) and is presented in Fig.9.

C. Case(iii): Stiction and Dead band

When \( f_s > f_d \) the valve with high initial static friction exhibits a jumpy behaviour that is different from dead band.

Figure 5: Valve Behaviour In The Absence Of Stiction

Figure 6: Step Response Of The Valve In The Absence Of Stiction
When the valve starts to move, the friction force reduces abruptly from $f_s$ to $f_d$, the initial velocity is faster making $f_s$ equal to $f_d$ leading to jumpy behaviour as shown in fig.8. The corresponding step response is given in Fig.9.

D. Case(iv): High Stiction

Under high stiction, the actuator piston applies increasing pressure in the air cylinder causes a temporary valve stem stopping and leads to a jumpy movement as shown in fig.10.

The valve response under this condition for a periodic ramp signal is shown in Fig.11. Under such high level of stiction, the stiction compensators will not give satisfactory results and the control valve has to be replaced.

Since the model is directly based on the dynamics of the valve, it is very simple to use for the purpose of simulation and can quantify stiction as a span of input signal. Also the parameters used in this model are easy to understand, realize and relate to real stiction behaviour. Though this
is an empirical model, it is observed that this model can correctly reproduce the behaviour of stiction model based on physical principles.

7. Effects of Valve Stiction under Closed Loop Condition

For assessment of closed loop behaviour, the valve output drives a first order plus dead time process $G_p(s)$ and receives its reference input from a PI controller $G_c(s)$ where,

$$G_p(s) = \frac{1.5e^{-5.93s}}{5.93s+1}, \quad G_c(s) = 1 + \frac{1.1}{2.5s}$$

Fig.12 shows the block diagram of a control loop in the presence of stiction. Valve dynamics are observed only after the starting of stem movement; stiction phenomenon if present will precede the valve dynamics. It is assumed that the valve is suffering from strong stiction. The triangular shape of the time trend of process output shown in Fig.13 is one of the characteristics of stiction [6] and this looks similar to Fig.2. The presence of stiction causes limit cycles of the process output. Fig.14 shows the mapping of controller output vs. valve position and it clearly shows the stiction phenomena in the valve. It is common practice to use the mapping of controller output vs. process output for valve diagnosis which is shown in Fig.15. However in this case such a mapping only shows elliptical loops with sharp turn around points. The reason is that, this map captures not only the valve characteristics but also the dynamics of the process $G_p(s)$, which in this case is a first order lag plus dead time.

This behaviour, results in oscillatory control signal and process output. The corresponding simulation results are shown in fig. 13. The above simulation results show that a high value of stiction leads to a high magnitude and high frequency oscillations. This result clearly exhibits the typical rectangular waveforms of the oscillations of the valve output and the triangular wave forms of the process output, which is the effect of stiction. The period of oscillation depends on the process dynamics, controller dynamics and valve characteristics. The response between controller output and valve position is shown in Fig.14. Here the value of slip jump(J) is much larger. In practical situation this response is difficult to obtain, since the valve position cannot be
measured directly. Usually the response between controller output and process output is obtained and is shown in Fig.15.

8. PRACTICAL EXAMPLES OF VALVE STICKTION

The effects of stiction from the investigation of data acquired by conducting an experiment on the laboratory flow control loop are explained in this section. The flow loop is a slave loop, cascaded with a master temperature control loop. In total 1100 samples are collected at a sampling rate of 0.01 sec. Fig.16 shows the step response of flow process showing stiction phenomena. Fig.17 shows the controller output Vs process output plot of laboratory flow loop which is similar to that of the simulated response shown in Fig.15 and that the model is validated.

9. CONCLUSION

In this work, a structurally simple and logically straightforward approach for modelling a valve stiction is proposed. The model has parameters that can be directly related to plant data and it produces the same behaviour as the physical model. The model needs only the specification of static and dynamic friction values. It overcomes the disadvantages of physical modelling of a control valve, which require the knowledge of the mass of the moving parts of the actuator, spring constant and the friction forces. The effect of the change of these parameters cannot easily be determined analytically. The proposed model overcomes some of the disadvantage of the existing models and this control valve stiction model can be used to study the stiction phenomena and its effects on closed loop control performance. Both closed loop and open loop results have been presented and validated to show the capability of the model. This data driven model is capable of handling stochastic inputs and can be used to perform simulation of stiction in Matlab's Simulink environment in the studies of stiction related control loop problems.

REFERENCES


**Author’s Biography**

Mrs. S.Sivagamasundari, currently working as a Reader in the Department of Electronics & Instrumentation Engineering, Annamalai University, India has a teaching experience of more than 17 years. She obtained her B.E. (Electronic and Instrumentation) degree and went on to complete her M.E. (Process control and Instrumentation), both from Annamalai University. Her research interests include Nonlinear Process Control and Instrumentation, Adaptive control and System identification. The author has also presented her research findings as papers in National and International conferences.

Dr. D.Sivakumar, a Professor in the Department of Electronics & Instrumentation Engineering, Annamalai University, India has a teaching experience of more than 25 years. He obtained both B.E. (Electronics & Instrumentation) and M.E. (Power Systems) degrees from Annamalai University. Further, he completed his Ph.D. Research programme at PSG College of Technology, Bharathiar University, Coimbatore. His research interests include Control System, Process Control, Fault Tolerant Control, Signal & Image Processing, Adaptive control and System identification. He has successfully guided two PhD scholars. He has to his credit research papers presented in many National and International conferences and published in National as well as International Journals. He is a life member in professional bodies like ISTE and System Society of India.
Investigation of Image Fusion using Curvelet Transform

Mary Praveena, S1, Radhika S, 2 and Ila Vennila3

1Research Scholar, Sri Ramakrishna Institute of Technology, Coimbatore - 641010, Tamil nadu, India.
2Student, ME-Applied Electronics, PSG College of Technology, Coimbatore - 641004, Tamil nadu, India.
3Supervisor, EEE Department, PSG College of Technology, Coimbatore 641004, Tamil nadu, India.
E-mail: praveena_infant@yahoo.co.in, radhisjr@gmail.com, iven@eee.psgtech.ac.in

ABSTRACT

Image fusion is the process of extracting meaningful visual information from two or more images and combining them to form a new image. In medical image fusion, multiple medical images, such as computed tomography (CT) and magnetic resonance (MR) images are fused into a new image to improve the information content for diagnoses. Some attempts have been proposed for the fusion of MR and CT images using the wavelet transform. Since medical images have several objects and curved structures, it is expected that the curvelet transform would be better in their fusion. Global Energy Merging scheme is a region-based analysis approach comparing to all other methods this method produces fused image having very high Edge and curve details.

Keyword: Curvelet Transform, Peak Signal To Noise Ratio, Tiling, Ridgelet Transform, Root Mean Square Error, Subband Filtering.

1. INTRODUCTION

Image fusion is important in many different image Processing fields such as satellite imaging, remote sensing and medical imaging. The study in the field of image fusion has evolved as a server for the advance in satellite imaging and then it has been extended to the field of medical imaging. With the development of multi-sensors, it is possible to obtain data from different sensors. Because of the different properties of multi-sensors, these images might provide totally different information. Several fusion algorithms have been proposed extending from the simple averaging to the curvelet transform.

Algorithms such as the intensity-hue-saturation (IHS) and the wavelet transform have proved to be successful in satellite image fusion. IHS belongs to colour image fusion algorithms. The wavelet transform for image fusion has also succeeded in both satellite and medical image fusion applications.

The basic limitation in the wavelet transform is in the fusion of curved objects. So the application of the curvelet transform for curved object image fusion would result in better fusion techniques. MR and CT imaging are of main concern for diagnostic purposes efficiency. A few attempts for curvelet fusion have been made in the field of satellite image fusion but no attempts in medical image fusion have been made using the curvelet transform.

The main objective of medical imaging is to obtain a high resolution image with as much details as possible
for the sake of diagnosis. It is known that there are several medical imaging

Both techniques give special sophisticated characteristics of the organ to be imaged. So, it is expected that fusion of MR and CT images of the same organ would result in an integrated image of much more details. Researchers have made few attempts for the fusion of MR and CT images. Most of these attempts are directed towards the application of the wavelet transform for this purpose. Due to the limited ability of the wavelet transform to deal with images with curved structures, the application of curvelet image fusion is proposed.

The curvelet transform is based on the segmentation of the whole image into small overlapping tiles and then the Ridgelet transform is applied to each tile. The purpose of the segmentation process is to approximate curved lines by small straight lines. The overlapping of tiles aims at avoiding edge effects. The Ridgelet transform itself is a 1-D wavelet transform applied in the Radon domain on each tile. The Radon transform is mainly presented as a tool for shape detection, especially of curved objects. Mainly the Curvelet transform was proposed for image denoising.

The paper is organized as follows. In section II, a new energy fusion method is proposed. Some relative equations and the detailed algorithm are given about this method.

2. GLOBAL ENERGY METHOD BASED ON REGION

The curvelet transform is based on the segmentation of the whole image. The low frequency band at each level can be segmented. Then Global Energy Method is performed. After all the processes above, a final fused image will be obtained by applying Inverse curvelet transform to the acquired coefficient matrix.

3. WAVELET FUSION

The most common form of transform type image fusion is the wavelet transform fusion due to its simplicity and its ability to preserve the time and frequency details of the image. A schematic diagram for the fusion of two images using the wavelet transform is depicted in Fig.(1). It can be defined considering the wavelet transform ú of two registered input images \( I_1(x, y) \) and \( I_2(x, y) \) together with the fusion rule \( ð \).

There are several wavelet fusion rules that can be used for selecting the wavelet coefficients. The most frequently used rule is the maximum frequency rule which selects the maximum coefficients from the wavelet transformed images. Then, the inverse wavelet transform \( ú^{-1} \) is computed, and the fused image \( I(x, y) \) is reconstructed:

\[
I(x, y) = ú^{-1} (ð (ú (I(x, y)), ú (I(x, y))))
\]

The wavelet transform concentrates on representing the image in multiscales and it's appropriate to represent linear edges. For curved edges, the accuracy of edge localization in the wavelet transform is small. So, there is a need for an alternative approach with high accuracy of curve localization such as the curvelet transform.

4. THE CURVELET TRANSFORM

The curvelet transform has evolved as a tool for the representation of curved shapes in graphical applications. Then, it was extended to the fields of edge detection and image denoising. Recently, some authors have proposed its application in image fusion. The algorithm of the curvelet transform can be summarized in the following steps:

A) The image is split up into three subbands using the additive wavelet transform.

B) Tiling is performed on subbands 1 Å and 2 Å.
C) The discrete Ridgelet transform is performed on each tile of the subbands \( \tilde{A}_1 \) and \( \tilde{A}_2 \).

A schematic diagram of the curvelet transform is depicted in Fig.(2)

A detailed description of these steps is presented below.

A. Subband Filtering

The purpose of this step is to decompose the image into additive components each of which is a subband of the image. This step isolates the different frequency components of the image into different planes without down sampling as in the traditional wavelet transform.

The "a trous" Algorithm given below is implemented for this purpose.

Given an image \( P \), it is possible to construct the sequence of approximations:

\[
f_1(P) = P_1, f_2(P_1) = P_2, f_3(P_2) = P_3, \ldots f_n(P_{n-1}) = P_n \quad (2)
\]

To construct this sequence, successive convolutions with a certain low pass kernel are performed. The functions \( f_1, f_2 \) and \( f_3 \) mean convolution with this kernel.

The wavelet planes are computed as the differences between two consecutive approximations \( P_{l-1} \) and \( P_l \).

\[
w_l = P_{l-1} - P_l \quad (3)
\]

where \( P_0 = P \) is the original image.

Thus, the reconstruction formula is given by:

\[
P = \sum_{l=1}^{n} W_l + P_r \quad (4)
\]

where \( P_r \) is the residual image.

B. Tiling

Tiling is the process by which the image is divided into overlapping tiles. These tiles are small in dimensions to transform curved lines into small straight lines in the subbands \( \tilde{A}_1 \) and \( \tilde{A}_2 \). The tiling improves the ability of the curvelet transform to handle curved edges. The purpose of the segmentation process is to approximate curved lines by small straight lines. The overlapping of tiles aims at avoiding edge effects.

C. Ridgelet Transform

The Ridgelet transform belongs to the family of discrete transforms employing basis functions. It is complicated to some extent. To facilitate its representation mathematically, it can be viewed as a wavelet analysis in the Radon domain. The Radon transform itself is a tool for shape detection. So, the Ridgelet transform was primarily a tool for ridge detection or shape detection of the objects in an image.

The Ridgelet basis function is given by:

\[
\phi(a,b,\theta) = a^{-\frac{1}{2}} \alpha \frac{x_1 \cos \theta + x_2 \sin \theta - b}{\lambda} \quad (5)
\]

Thus, the Ridgelet coefficients are represented by:

\[
R(a,b,\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) \phi(a, b, \theta)(x_1, x_2) dx_1 dx_2 \quad (6)
\]

This transform is invertible and the reconstruction formula is given by:

The Radon transform for an object \( f \) is given by:

\[
R(\theta) = \int_{-\infty}^{\infty} f(x_1, x_2) \delta(x_1 \cos \theta + x_2 \sin \theta - t) dx_1 dx_2 \quad (7)
\]

Thus, the Ridgelet transform can be represented as follow:

\[
R(a,b,\theta) = \int_{-\infty}^{\infty} \tilde{R}(\theta, t) a^{-\frac{1}{2}} \lambda \left( \frac{x_1 \cos \theta + x_2 \sin \theta - t}{\lambda} \right) dt \quad (8)
\]
Hence, the Ridgelet transform is the application of a one-dimensional wavelet transform to the slices of the Radon transform where the angular variable $\theta$ is constant and $t$ is varying. A schematic diagram of the Ridgelet transform is shown in Fig. (3). To make the Ridgelet transform discrete both the Radon transform and the wavelet transform have to be discrete.

5. THE PROPOSED FUSION ALGORITHM

It is known that different imaging modalities are employed to depict different anatomical morphologies. CT images are mainly employed to visualize dense structures such as bones. So, they give a general shape of objects and fewer details. On the other hand, MR images are used to depict the morphology of soft tissues. So, they are rich in details. Since these two modalities are of a complementary nature, the objective is to merge both images in a single image to obtain as much information as possible.

A Curvelet based approach is introduced for this purpose. It is summarized as follows:

1. The MR and CT images are registered.
2. The curvelet transform steps are performed on both the MR and CT images.
3. The maximum frequency fusion rule is used for the fusion of the Ridgelet transforms of the subbands $A_1$ and $A_2$ for both MR and CT images.
4. An inverse curvelet transform is performed on MRI image and the fused subbands $A_1$ and $A_2$.

These steps are expected to merge the details in both images into a single image with much more details.

GEM algorithm is used for different levels of decompositions.

**GEM Algorithm steps**

The following algorithm performs the fusion process:

1. Initialization: define the area size $A$ which will be used around each location $p$; $n=1$.
2. Input: image $F_1$ and $F_2$, set $X=F_1$ and $Y=F_2$
3. While $n<N$ do $n=n+1$; $X \rightarrow$ DWT $\rightarrow D_L L_n$, $D_L H_n$, $D_H H_n$; set $X= D_L L_n$, $Y= D_L L_n$ and go to 3
4. Image Tiling
5. Ridgelet transform
6. Fused image: Select for each location $p$ at $(x, y)$ in the transformed tiles compute the following region energy values using the equation

$$E_{ij}(P) = \sum_{p \in A} W(P) D_{ij}(p)^2$$

Where

$E_{ij}(P); E_{ij}(P)$ with $ij = LL, LH, HL, HH$

Merge: firstly low frequency band is calculated

$$e_{LL}(P) = \max(E_{LL}(P), E_{LL}(P))$$

$$D_{LL}(P) = D_{LL}(P)$$ IF $e_{LL}(p)$ come from $X$

else $D_{LL}(P) = D_{LL}(P)$

7. Then high frequency bands are calculated:

When $M_{ij} (p) \geq \alpha$

Do

$$D_{ij}(P) = W_{ij}(P) \ast D_{ij}(P) \ast W_{ij}(P) \ast D_{ij}(P)$$

When $M_{ij} (p) < \alpha$

Do

$$e_{ij}(P) = \max(E_{LL}(P), E_{LL}(P))$$

$$D_{ij}(P) = D_{ij}(P)$$ IF $e_{LL}(p)$ come from $X$

else $D_{ij}(P) = D_{ij}(P)$
Investigation of Image Fusion using Curvelet Transform

8. While \( n \neq -1 \)
Apply Inverse Curvelet Transform to transformed tiles. 
Obtaining \( D^{(n+1)W} \);
\( n\leftarrow n-1 \) and go to 8

9. Output: fused image \( Z = D^{W} \)

6. PERFORMANCE MEASURES

a. Root Mean Square Error (RMSE)
As a quality measure, the RMSE (Root Mean Square Error) is used. It is expressed as follows:

\[
RMSE = \left( \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} (x_{R}(n,m) - x_{F}(n,m))^2 \right)^{1/2}
\]

Where \( x_{R} \) is the ideal reference, \( x_{F} \) the obtained fused image, and \( M, N \) are the dimensions of the images. Root mean square error indicates how much error the fused image \( x_{F} \) conveys about the reference \( x_{R} \).

Thus, the lower the RMSE between \( x_{F} \) and \( x_{R} \), the more likely resembles the ideal \( x_{R} \).

The mean-square error function is commonly used because:

- It is easy to compute,
- It is differentiable implying that a minimum can be sought,
- It corresponds to “signal energy” in the total error, and
- It has nice properties \( \text{vis à vis} \) Parseval’s theorem.

b. Peak Signal To Noise Ratio (PSNR)

The PSNR is defined as follows:

\[
PSNR = 10 \log_{10} \left( \frac{f_{\max}^2}{\text{RMSE}^2} \right)
\]

where \( f_{\max} \) is the maximum gray scale value of the pixels in the fused image. The higher the value of the PSNR, the better the fusion performance. The RMSE between the fusion result and both the MR and CT images is estimated and two values of PSNR for both curvelet and wavelet fusion results are obtained.

7. EXPERIMENTAL RESULTS

CT and MR scans of the brain are used as input images shown in Figs. (4) and (5), respectively. The wavelet fusion result is given in Fig. (6) And the curvelet fusion result is given in Fig. (7). From the fusion results of figs. (6) and (7), It is clear that the curvelet fusion result has a better visual quality than the wavelet fusion result.

These values reflect the ability of the curvelet transform to capture features from both the MR and CT images. From these results, it is clear that the proposed algorithm has succeeded in obtaining better results than the wavelet transform from both the visual quality and PSNR points of view.

8. CONCLUSIONS

The paper has presented a new trend in the fusion of MR and CT images using the curvelet transform. A comparison study has been made between the traditional wavelet fusion and the proposed curvelet fusion. The experimental study shows that the application of the curvelet transform in this fusion is superior to the application of the traditional wavelet transform. The obtained curvelet fusion results have higher PSNR values.
than the wavelet fusion results. Also curved visual details are better in the curvelet fusion results than in the wavelet fusion results.

REFERENCES


Investigation of Image Fusion using Curvelet Transform


[16] (Computed Tomography (CT) Scan) http://www.uhs.net/uma/CT/20Scan.pdf


Author’s Biography

Ms. Mary Praveena is working as a Lecturer in Sri Ramakrishna Institute of Technology, Coimbatore and pursuing PhD under the Guidance of Dr.Ila Vennila. Her research interest includes Digital Image Processing, Neural Networks, and Genetic Algorithm. Her publication includes several papers in National / International conferences.


Dr. IlaVennila is working as an Assistant Professor in Electrical and Electronics Department, PSG college of Technology, Coimbatore. Her research areas of interest include Digital Signal Processing and Digital Image Processing. Dr.Vennila has over 23 years of Experience in teaching. Her publication includes 10 journal papers and 30 National / International Conferences.