<table>
<thead>
<tr>
<th>CONTENTS</th>
<th>PAGE NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. An Effective Secret Sharing Scheme for N Out Of N Scheme Using Modified Visual Cryptography A. Sreekumar and S. Babusundar</td>
<td>1692</td>
</tr>
<tr>
<td>2. Associative Text Classification for Online Database V. Srividhya, R. Anitha</td>
<td>1700</td>
</tr>
<tr>
<td>3. A Novel Algorithm For Frequent Item – Sets Mining Using Reduced Database Scan Approach Mahesh. H. Panchal, Bhagirath. P. Prajapati</td>
<td>1707</td>
</tr>
<tr>
<td>4. A Novel Approach For Content Based Image Retrieval Srinivasa Kumar Devireddy</td>
<td>1716</td>
</tr>
<tr>
<td>6. Detection, Counting And Classification Of Moving Objects By Using Real Time Traffic Flux Through Differential And Rule Based Analysis H. S. Mohana and M. Aswathakumar</td>
<td>1741</td>
</tr>
<tr>
<td>10. Implementation of FPGA Based Incremental PID Controller Using Conventional Method and Distributed Arithmetic Algorithm Mariamma John, B. Thilagavathi, Anitha Mary. X</td>
<td>1776</td>
</tr>
</tbody>
</table>
An Effective Secret Sharing Scheme for N out of N Scheme Using Modified Visual Cryptography
A. Sreekumar¹ and S. Babusundar²

ABSTRACT
Secret sharing is concerned with the problem of how to distribute a secret among a group of n participating individuals, or entities, so that only pre-designated collections of individuals are able to recreate the secret by collectively combining their shares of secret. Sharing schemes are useful in military and civilian applications. In the traditional Secret Sharing Schemes, a shared secret information cannot be revealed without any cryptographic computations. Various Secret Sharing Schemes have been proposed. However, the size of the shares and implementation complexity in these schemes depend on the number of participants. In other words, when a great number of participants are involved, the scheme will become impractical. A secret sharing scheme is called efficient if the total length of the n shares is polynomial in n: In the traditional Visual Secret Sharing Schemes, a shared secret information can be revealed without any cryptographic computations. In this paper we propose an n out of n uniform secret sharing scheme based on visual cryptography. This scheme provides an efficient way to hide a secret information in different shares. Further more, the size of the shares is just 1 bit more than the size of the secret, and so it does not vary when the number of participants differs.

1. INTRODUCTION
Secret sharing scheme is a method of sharing a secret information among a group of participants. In a secret sharing scheme, each participant gets a piece of secret information, called a share. When the allowed coalitions of participants pool their shares, they can recover the shared secret; on the other hand, any other subsets, namely non-allowed coalitions, cannot recover the secret information by pooling their shares. The collection of subsets of participants that can reconstruct the secret in this way is called access structure. Secret Sharing was introduced by Blakley [8] and Shamir [1] in 1979. Shamir’s solution is based on the property of polynomial interpolation in finite fields; Blakley formulated and solved the problem in terms of finite geometries. The first secret sharing schemes considered were threshold schemes. A (k, n) threshold scheme allows a secret to be shared among n participants in such a way that any k of them can recover the secret, but any k-1, or fewer, have absolutely no information on the secret.

Asmuth and Bloom [2] implemented a (k, n) threshold scheme based on Chinese Remainder Theorem in 1983. In [25]D. R. Stinson and S. A. Vanstone introduced anonymous threshold scheme. Informally, in an anonymous secret sharing scheme the secret is reconstructed without the knowledge of which participants hold which shares. In such schemes the computation of the secret can be carried out by giving the shares to a black box that does not know the identities of the participants holding those shares. During 1987 Ito,
Saito, and Nishizeki [16] described a generalized method of secret sharing scheme whereby a secret can be divided among a set \( P \) of trustees such that any qualified subset of \( P \) can reconstruct the secret and unqualified subsets cannot. Phillips and Phillips [22] considered a different model for anonymous secret sharing schemes. In their model, different participants are allowed to receive the same shares. Further results on this type of anonymous secret sharing schemes can be found in [10].

Redistributing secret shares to new access structures has been considered in [9]. Secret Sharing schemes based on Chinese Remainder Theorem is introduced by Mignotte [20]. D. R. Stinson [26] gives a comprehensive introduction to this topic.

A black-box secret sharing scheme for the threshold access structure is one which works over any finite Abelian group. G. Bertilsson and Ingemarsson [7] describes a construction method of practical secret sharing schemes using Linear Block Codes.

A more general approach has been considered by Karnin, Greene and Hellman [17] who invented the analysis (limited to threshold scheme) of secret sharing schemes when arbitrary probability distributions are involved.

Some other general techniques handling arbitrary access structures are given by Simmons, Jackson, and Martin [19] [24] and also by suggested by Kothari [18].

In [11] Brickell introduced the vector space construction which provides secret sharing schemes for a wide family of access structures. In [26] Stinson proved that threshold schemes are vector space access structures. Various Secret sharing schemes were proposed, but most of them need a lot of computations to decode the shared secret information. While in threshold schemes proposed by Blakley [8] and Shamir [23] and in the vector space schemes given by Brickell [11] the shares have the same size as the secret, in the schemes constructed by M. Ito, A. Saito, and T. Nishizeki [16] for general access structures the shares are, in general, much larger than the secret.

Subsequently, Benaloh and Leichter [6] gave a simpler and more efficient way to realize such schemes. They also proved that no threshold scheme is sufficient to realize secret sharing on general monotone access structures. In support of their claim, they have shown that there is no threshold scheme such that the access structure \((A \lor B) \land (C \lor D)\) can be achieved.

In [5] Benaloh describes a homomorphism property that is present in many threshold schemes which allows shares of multiple secrets to be combined to form “composite shares” which are shares of a composition of the secrets. An important issue in the implementation of secret sharing schemes is the size of shares, since the security of a system degrades as the amount of the information that must be kept secret increases. If one requires that non-qualified set of participants should have no information on the secret, then the size of the shares cannot be less than the size of the secret. This fact is established by E. D. Karnin, J. W. Greene and M. E. Hellman [17]. In [6] J. C. Benaloh and J. Leichter, proved that there exists an access structure (namely the path of length three) for which any secret sharing scheme must give to some participant a share which is from a domain larger than that of the secret.

Capocelli, De Santis, Gargano and Vaccaro [12] proved that there exist access structures for which the best achievable information rate (i.e., the ratio between the size of the secret and that of the largest share) is bounded away from 1.

Tompa and Woll [27] considered the issue of cheaters in 1988 and could able to detect cheaters. A cheater might
tamper with the content of a share and make the share unusable for combining to retrieve the secret.

The problem of identifying the cheater is solved by the authors. In a sense, it is an improvement on the works of Shamir [23]. In 1994, Naor and Shamir [21] invented a new type of Secret sharing scheme called visual cryptography scheme. It could decode the secret (printed text, hand written notes, pictures, etc.) directly without performing any computation, and the decoder of this scheme was the human visual system. For example, in a \((k, n)\) visual cryptographic scheme, a dealer encodes a secret into \(n\) shares and gives each participant a share, where each share is a transparency. The secret is visible if \(k\) (or more) of participants stack their transparencies together, but none can see the shared secret if fewer than \(k\) transparencies are stacked together.

Until the year 1997, although the transparencies could be stacked to recover the secret image without any computation, the revealed secret images (as in [3] [4] [14] [21]) were all black and white. In [28], Verheul and Van Tilborg used the concept of arcs to construct a colored visual cryptography scheme, where users could share colored secret images. The key concept for a \(c\)-colorful visual cryptography scheme is to transform one pixel to \(b\) sub-pixels, and each sub-pixel is divided into \(c\) color regions. In each sub-pixel, there is exactly one color region colored, and all the other color regions are black. The color of one pixel depends on the interrelations between the stacked sub-pixels. For example, if we want to encrypt a pixel of color \(c_i\), we color region \(i\) with color \(c_i\) on all sub-pixels. If all sub-pixels are colored in the same way, we sees color \(c_i\), when looking at this pixel; otherwise one sees black.

A major disadvantage of this scheme is that the number of colors and the number of sub-pixels determine the resolution of the revealed secret image. If the number of colors is large, coloring the sub-pixels will become a very difficult task, even though we can use a special image editing package to color these sub-pixels. How to stack these transparencies correctly and precisely by human beings is also a difficult problem. Another problem is that when the number of sub-pixels is \(h\), the loss in resolution from the original secret image to the revealed image becomes \(h\).

In [15], Hwang proposed a new visual cryptography scheme which improved the visual effect of the shares (the shares in their scheme were significant images, while those in the previous scheme were meaningless images). Hwang’s scheme is very useful when we need to manage a lot of transparencies; nevertheless, it can only be used in black and white images. For this reason, Chang, Tsai and Chen proposed a new secret color image sharing scheme [13] based on modified visual cryptography.

A major disadvantage of this scheme is that the size of the share is in proportion with the number of participants, i.e., the more the participants, the larger the share will become. The ratio of the size of one share to the size of the secret is called the information rate.

2. PECULIARITY OF EVEN PARITY STRINGS

Any information can be encoded as a binary string. So it is sufficient to consider only binary strings in any secret sharing schemes. The proposed scheme is based on the following theorem:

**Theorem 2.0.1** Let \(T\) be an even parity binary string of length \(t\). Then we can find two POB-Numbers \(A\) and \(B\), both \(\in \text{POB}(t, \left[ \frac{t}{2} \right])\) such that \(T = A \oplus B\).

**Proof**: We can assume, without loss of generality that, the leading \(2m\), \((0 \leq m \leq \left[ \frac{t}{2} \right])\) digits of \(T\) are 1s and remaining \(t - 2m\) (\(\geq 0\)) digits are 0s. Now, let \(A = PQ\) be the binary string obtained by concatenating the strings \(P\).
An Effective Secret Sharing Scheme for N out of N Scheme Using Modified Visual Cryptography

and Q, where \( P \) is the string having exactly \( m \) 1s, followed by \( m \) 0s, and \( Q \) is the string having exactly \( \frac{1-2m}{2} \) 1s and \( \frac{1-2m}{2} \) 0s. Then the choice \( B = PQ \), where, \( P \) is the Boolean complement of \( P \), will prove the theorem. However, such a decomposition, in general, need not be unique. We can see that both \( A \) and \( B \in \text{POB}(t, \frac{1}{2}) \) number system. Also, once we find \( A \), we can get \( B \) at ones, \( B = \oplus A \). It may be noted that, among the \( 2m \) 1s in \( T \), exactly \( m \) 1s are in matched position with \( P \), and the other \( m \) 1s are in matched position with \( Q \). The bits in \( P \) and \( Q \), corresponding to a 0 in \( T \) are same (either both 0 or both 1), they are assigned randomly, with ensuring the only condition that, they \( \in \text{POB}(t, \frac{1}{2}) \).

2.1. The Proposed Secret Sharing Scheme

In this section, we present our method to construct an \( n \) out of \( n \) secret sharing scheme based on the modified visual cryptography. Assume that the secret can be represented as a binary string \( b_1b_2b_3 \ldots b_t \). Our scheme will generate \( n \) shares after attaching a single bit, \( b_{t+1} \) at the end of the secret. The resulting structure of the share can be described as an \( n \times (t+1) \) Boolean matrix \( S = [S_{ij}] \), where \( 1 \leq i \leq n; 1 \leq j \leq (t + 1) \). The original secret will be revealed by performing the “XOR” operation (denoted by \( \oplus \) and read as ring sum) on each row in \( S \), and deleting the last bit attached at the end. For an \( n \) out of \( n \) secret sharing scheme, the construction can be described by any Boolean string \( C \). The construction is considered valid if, for any Boolean string \( S \in C \), the ring sum, \( \oplus \), of each row in \( S \) satisfies the following equation:

\[
b_j = S_1 \oplus S_2 \oplus S_3 \oplus \ldots \oplus S_n,
\]

for \( j = 1, 2, \ldots, t \).

Here we define \( S \) is a uniform code, if, each row of \( S \) is a uniform code.

2.2. A Uniform 2 Out of 2 Construction

We now describe the construction details of a uniform 2 out of 2 secret sharing scheme and extend it to a uniform \( n \) out of \( n \) scheme in the next section. Let \( B = b_1b_2b_3 \ldots b_t \), be the secret information to be shared between two participants. We describe an efficient (2, 2) scheme by making use of the theorem 2.0.1. First of all, the necessary condition to use the theorem is that, the concerned string must be even parity. So, we extend the secret by appending a single bit at the right end. If we discard the appended last bit, we get precisely the secret. The length of the extended string is just one more than that of the secret.

The Algorithm 1 extends the string and makes the resulting string an even parity.

Now, using construction method in theorem 2.0.1, we split this extended string and obtain the two shares. The very simple algorithm 2 nds the decomposition of the extended string, as in theorem 2.0.1. The algorithm 3 shares any binary string between two shares, by using algorithm 1 and then algorithm 2.

Recovery : From \( E_{n+1} = \sigma_{1+1}^{(1)} \oplus \sigma_{1+1}^{(2)} \), it follows, if we just discard last bit of \( E_{n+1} \), we get \( B \), i.e, the recovery procedure is that, just \( \oplus \) the two shares, we get the extended string, and discard the last appended bit we get the secret. hence the following lemma :

Lemma 1 The Algorithm 3 described below a (2,2)-modified visual cryptography scheme, in which the size of the secret is just one bit more than the size of secret. More over, all the shares are \( \text{POB}(t + 1, \frac{1+t}{2}) \)-numbers.

The two shares are constructed by using the Algorithm 3 described below :

Algorithm 1 [Append a single bit at the end] Input : A binary string \( B = b_1b_2 \ldots b_t \) of length \( t \).
Example 1:
Let the secret B be
10011 00101 00011 10010 00101 10100
By Step 1 of Algorithm 2, Initialize S1 and S2, null.
Step 2 of algorithm 2, S1 is computed as
1 * * 01 * * 0 * 1 *** 010 * * 1 * * 0 10 * 1 * * 0
(Here * null bits.) Step 3 of Algorithm 2, S1 is randomly set as
111010010010010011101001001110
Step 4 of Algorithm 2, S2 = S1 ⊕ Bt−1 =
011101000101001101010100101001
Recovery: Compute S1 ⊕ S2 and get
Bt = 10011001001101010100101001010101
Last bit is 1 and is deleted to get B : 10011 00101 00011 10010 00101 10100.
2.3. A Uniform n Out of n Construction
Algorithm 5: [Sharing a secret among n blocks]
Input: A binary string Bt = b1b2 ... bt of length t.
Output: n blocks S1, S2, ..., Sn of length t + 1.
Algorithm 6: [Recover the secret information]
Input: n shares S1, S2, ..., Sn of length t + 1.
Output: The secret information Bt = b1b2 ... bt.
Lemma 2 The Algorithm 5 described above, is an (n,n)-modified visual cryptography scheme, in which the size of the secret is just one bit more than the size of secret. More over, all the shares are POB(t + 1, | 1 + 1 |) - numbers.
extended string $B_{t+1}$. Note that the last bit appended is insignificant. In Step 2, it generate $n-2$ shares, $S_2, S_3, \ldots, S_{n-1}$. They are all random POB $(t+1, [\frac{t+1}{2}])$ numbers. In Step 3, from the equation,

$$K_{t+1} = B_{t+1} \oplus S_2 \oplus \ldots \oplus S_{n-1}$$

(2)

the following equation holds :

$$B_{t+1} = K_{t+1} \oplus S_2 \oplus \ldots \oplus S_{n-1}$$

(3)

In step 4, we ensure that $K_{t+1}$ is even parity. If not, the last insignificant bit will be toggled to make it even parity. It also toggles the last bit of $B_{t+1}$, so that equation (3) is still valid. Finally, in step 5, share $K_{t+1}$, between two shares $S_1 \oplus S_n$ by Algorithm 2 with input $K_{t+1}$.

Example 2:

For a (5, 5) threshold scheme, secret $B = 101101110$ is taken. By step 1, the extended string is, $B_{t+1}$ of length 10 is 10110111 00

Randomly assign 5 1s and 5 0s to 3 rows $\{S_2, S_3, S_4\}$ in $S$. Therefore,

$S_2 = 10100010101$, $S_3 = 0101010110$, and $S_4 = 1100101010$.

Step 3. Computes $K = 1001100101$, and In Step 5,

1001100101 0 is split into

$S_1 = 1010100100$, and $S_5 = 0011010110$.

All the 5 shares are listed below:

$S_1 = 1010100100$, and $S_5 = 0011010110$.

Recovery: Computes $S_1 \oplus S_2 \oplus S_3 \oplus S_4 \oplus S_5$ and get $B_{t+1} = 1011011101$.

Deleting the last bit of $B_{t+1}$, we get the secret as $B_t = 10110111 0$.

3. SECURITY ANALYSIS

In this section, we discuss the security of the proposed scheme. In order to show the security of the uniform 2 out of 2 construction, suppose an illegal user gets one of the two shares. Lemma 3 shows guessing the secret correctly is very difficult.

**Lemma 3** With only one share, the probability of guessing the shared secret correctly in a uniform construction is \(\left(\frac{t+1}{2}\right)^{-1}\).

**Proof** : In a uniform construction, it is easy to observe that each share contains \(\left(\frac{t+1}{2}\right)\) 1s. There are \(\left(\frac{t+1}{2}\right)\) variations for a block, and the probability of guessing one block correctly is \(\left(\frac{t+1}{2}\right)^{-1}\). Hence the probability of an illegal user, who has only one share, guessing the shared secret is \(\left(\frac{t+1}{2}\right)^{-1}\).

In order to show the security of an $n$ out of $n$ uniform construction, suppose there are fewer than $n$ participants cooperating to guess the shared secret. Lemma 4 shows that even though there are $n-1$ participants cooperating, the probability of guessing the shared secret correctly is still very low.

**Lemma 4** : The probability of guessing the shared secret correctly in a uniform construction is $\left(\frac{t+1}{2}\right)^{-1}$, if only $n-1$ shares are used to guess the share.

**Proof** : The proof is similar to that of Lemma 3.

**Conclusions**

We have presented a secret sharing scheme, in which the size of a share is just one bit more than the original secret size.
REFERENCES


An Effective Secret Sharing Scheme for N out of N Scheme Using Modified Visual Cryptography

1699


Author’s Biography

A Sreekumar received the M. Tech degree in Computer Science from I.I.T. Madras in the year 1992. Currently he is working as a Lecturer at Cochin University of Science and Technology. He had 16 years of teaching experience

Dr. Babusundar received the M. Tech and Ph.D Degrees from Cochin University and working as Professor of Computer Applications in CUSAT. He had 26 years of teaching experience and his area of interest is Fuzzy logic Applications, Artificial Intelligence, Multimedia and Cryptography.
Abstract
Text categorization is regaining interest with the prevalence of digital documents and the wide use of e-mail and web documents, and is becoming a central problem in digital text collections. There have been many approaches to solve this problem, mainly from the machine learning community. This paper explores the use of association rule mining in building a text categorizer. This approach has the advantage of a very fast training phase, less memory usage and the rules of the classifier generated are easy to understand. The investigation leads to conclude that association rule mining is a good and promising strategy for efficient automatic text categorization.

Keywords: Text Categorization, Text Mining, Association Rules and Classification.

1. Introduction
Amazing development of Internet and digital library has triggered a lot of research areas. Text categorization is one of them. Text categorization is a process that group text documents into one or more predefined categories based on their contents [1]. It has wide applications, such as email filtering, category classification for search engines and digital libraries.

Basically there are two stages involved in text categorization. Training stage and testing stage. In training stage, documents are preprocessed and are trained by a learning algorithm to generate classifier. In testing stage, a validation of classifier is performed. There are many traditional learning algorithms to train data. Examples include Decision trees, Naïve-Bayes (NB), Support Vector Machines, k-Nearest Neighbor (kNN), Neural Network (NN), etc. Nowadays, text categorization becomes fundamental given the large number of on-line documents that have to be sorted and grouped. For example large companies could use text classifiers for in-coming e-mail triage and memo categorization. Text classifiers can be used to classify web pages, in-coming emails, memos, news and any other text collection. Building a text classifier usually necessitates a training set consisting of a collection of text documents already associated with topical categories. Once a classifier is built with the training set, a test set, consisting of documents with known categories, is classified and the found class labels compared to the existing categories to determine the effectiveness of classifier.

This paper exploits the use of association rules mining in building categorization system from relatively large training set. The remainder of the paper is organized as follows: Section 2 gives an overview of related work in text categorization and association rule mining. Section 3 shows the new categorization approach. Experimental results are described in section 4. Summarization of research and discussion and research in future direction are narrated in Section 5.
2. RELATED WORK

Many text classifiers have been proposed in the literature using machine learning techniques, probabilistic models, etc. Although a lot of approaches have been proposed, automated text categorization is still major area of research. The use of association rule mining for building classification models is very new. This classification system discovers the strongest association rules in the database and uses them to build categorizer.

In the following subsections a more detailed overview of the related work is presented from both domains: text categorization and association rule mining

2.1 Text Categorization

In the past decade, great attention was paid to the text categorization problem. Most of the text classifiers that were developed and proposed are either machine learning based or statistical based. Classifiers based on probabilistic models have been proposed starting with the first presented in literature by Maron in 1961 and continuing with naïve Bayes [7] that proved to perform well. ID3 and C4.5 are well-known packages whose cores are making use of decision trees to build automatic classifiers [5, 6]. K-nearest neighbor (k-NN) is another technique used in text categorization [11]. Another method to construct a text categorization system is by an inductive learning method. This type of classifier is represented by as set of rules in disjunctive normal form that best cover the training set [4,8,9]. As reported in [12] the use of bigrams improved text categorization accuracy as opposed to unigrams use. In addition, in the last decade neural networks and support Vector Machines (SVM) were used in text categorization and they proved to be powerful tools [10,14].

2.2 Association Rule Mining

2.2.1 Association Rules Generation

Association rule mining has been extensively investigated in the data mining literature. Many efficient algorithms have been proposed. The most popular being is apriori [2] and FP-Tree growth [3]. Association rule mining typically aims at discovering associations between database items in a transactional. Given a set of transactions \( D = \{T_1, \ldots, T_n\} \) and a set of items \( I= \{i_1, \ldots, i_m\} \) such that any transaction \( T \) in \( D \) is a set of items in \( I \), an association rule is an implication \( A \rightarrow B \) where the antecedent \( A \) and the consequent \( B \) are subsets of a transaction \( T \) in \( D \), and \( A \) and \( B \) have no common items. For the association rule to be acceptable, the conditional probability of \( B \) given \( A \) has to be higher than a threshold called minimum confidence. Association rules mining is normally a two-step process, wherein the first step frequent item sets are discovered (i.e., item-sets whose support is no less than a minimum support) and in the second step association rules are derived from the frequent item-sets.

2.2.2 Associative Classifiers

Besides the classification methods, associative text categorization, and a new method that builds associative general classifiers. In this case the association rule mining represents the learning method. The main idea behind this approach is to discover strong patterns that are associated with class labels. The next step is to take advantage of these patterns such that a classifier is built and new objects are categorized in the proper classes.

3. BUILDING AN ASSOCIATIVE TEXT CLASSIFIER

In this paper, a method to build a categorization system that merges association rule mining task with the classification problem is presented. Given a data collection, a number of steps are followed until the
classification model is found. Data preprocessing represents the first step. The next step in building the associative classifier is the generation of association rules using FP-growth algorithm. The last stage in this process is represented by the use of the association rules set in the prediction of classes for new documents.

3.1 Association Rule Generation: The FP-Growth Algorithm

The FP-growth algorithm is currently one of the fastest approaches to frequent item set mining. The association rules discovered in this stage of the process is further processed to build the associative classifier.

The main bottleneck of the Apriori-like methods is at the candidate set generation and test. This problem was dealt with by introducing a novel, compact data structure, called Frequent Pattern tree, or FP-tree then based on this structure, an FP-tree-based pattern fragment growth method was developed, FP-growth. Figure 1 presents the pseudo code of FP-Growth algorithm.

The basic principle of FP-Growth is to work in a divide and conquer manner. Compared to Apriori, the FP-Growth algorithm requires only two scans on the database. It first computes a list of frequent items sorted by frequency in descending order (F-List) during its first database scan. In its second scan, the database is compressed into a FP-tree. Then FP-Growth starts to mine the FP-tree for each item whose support is larger than \( \xi \) by recursively building its conditional FP-tree. The algorithm performs mining recursively on FP-tree. The problem of finding frequent itemsets is converted to searching and constructing trees recursively.

3.2 Classification Process

Given a collection of documents, the first step is to index them to produce document representations. In the full text logical view, a representation of a document \( d_j \) is the set of all its terms (or words). Each term of the document representation is considered as a separate variable or feature. Bag-of-words technique is used for this purpose. From this, a subset of the terms to represent the documents is selected, through a process called feature selection, to reduce the document representations. For this purpose, preprocessing techniques like stop-word elimination, stemming and TF-IDF are used. Then the FP-Growth Algorithm is used to generate a set of association rules, which are used during the classification process.

| Input : database DB, minimum support |
| Output : the complete set of frequent patterns |
| Method : FP-Growth(DB, \( \xi \)) |
| Define and clear F-List : F[] : |
| for each Document Ti in DB do |
| for each Term ai in Ti do |
| F[ai] ++; |
| end |
| end |
| Sort F[]; |
| Define and clear the root of FP-tree : \( r_i \); |
| for each Document Ti in DB do |
| Make Ti ordered according to F; |
| Call ConstructTree(Ti; r); |
| end |
| for each term ai in I do |
| Call Growth(r, ai, \( \xi \)); |
| end |

Figure 1: FP-Growth Algorithm

To improve the effectiveness of the classification process, the classifier goes through a process of fine tuning its internal parameters. This is accomplished using a training set. Once its parameters have been fine tuned, the classifier is used to classify the new documents. This process is discussed in the next section. The final phase
of the process is to evaluate the effectiveness of the classification. The evaluation of a classifier is done by comparing the results with standard machine learning classifiers. This process is pictorially given in Fig. 2.

3.3 Prediction of Classes Associated with New Documents

The set of rules that were selected represent the actual classifier. This categorizer will be used to predict which class a new document will be attached. Given a new document, the classification process searches in this set of rules for finding those classes that are the closest to be attached with the document presented for categorization. This subsection discusses the approach for labeling new documents based on the set of association rules that forms the classifier.

Given a document to classify, the terms in the document would yield a list of applicable rules. If the applicable rules are grouped by category in their consequent part and the groups are ordered by the sum of rules confidences, the ordered groups would indicate the most significant categories that should be attached to the document to be classified. This order category is named as dominance factor $\delta$. The dominance factor allows us to select among the candidate categories only the most significant. When $\delta$ is set to a certain percentage a threshold is computed as the sum of rules confidences for the most dominate category times the value of the dominance factor. Then, only those categories that exceed this threshold are selected. The function (TakeKClasses) selects the most $k$ significant classes in the classification algorithm. The algorithm used is given in Fig. 3.

**Algorithm**: Classification of a new object

**Input**: A new object to be classified $o$; the Associative classifier (ARC); the dominance factor $\delta$; the confidence threshold $T$

**Output**: Categories attached to the new object

1. $s \gets \emptyset$
   /*set of rules that match*/
2. foreach rule $r$ in ARC (the sorted set of rules
3. if ($r \subset o$) {count++}
4. if (count == 1)
5. fi.conf $\gets r.conf$ /*keep the first rule confidence*/
6. $S \leftarrow S \cup r$
7. else if ($r.conf > fi.conf - T$)
8. $S \leftarrow S \cup r$
9. else
10. exit
11. divide $S$ in subsets by category: $S_1, S_2, \ldots, S_n$
12. foreach subset $S_1, S_2, \ldots, S_n$

Figure 2: The Text Classification Process Used In This Paper
(12) sum the confidences of rules and divide by the number of rules in $S_k$

(13) if it is single class classification

(14) put the new document in the class that has the highest confidence sum

(15) else /*multi-class classification*/

(16) TakeKClasses($S, \delta$)

(17) assign these $k$ classes to the new document

Figure 3: Classification of a New Object

4. EXPERIMENTAL RESULT

4.1 Experiment Data

The Reuters 21578 text collection was used as benchmarks in evaluating the system. Reuters 21578 is split into two parts: one part for training and a second part for testing. The ModApte version of split is used in this research. This split leads corpus of 12,202 documents consisting of 9,603 training documents and 3,299 testing documents. There are 135 topics to which documents are assigned [15]. Finally, the classifier was tested with ten most populated categories with largest number of documents assigned to them in training set. As most of the researches [13] have following this strategy, the present work uses the same approach for evaluation so that the results can be compared with the standard techniques. By retraining only the ten most populated categories, there are total of 6488 training documents and 2425 testing documents.

4.2 Experimental Results

When dealing with multiple classes there are two possible ways of averaging these measures, namely macro average and micro average. In micro average for all classes, an average of all classes is computed and the performance measure obtained there in. The macro-average weights equally all the classes, regardless of how many documents belonging to it. The micro average weights equally all the documents, thus favoring the performance of all classes.

Table 1 shows a comparison between the proposed ARC-FG (Association Rule based categorizer - Frequent Growth) classifier and other well-known methods. The measures used are precision/recall-breakeven point; micro average and macro average on ten most populous Reuters categories. The proposed system proves to perform well as compared to the other methods. In general, the functioning of the proposed algorithm is in par with the existing state of the art text classifiers. In addition to these results, the system has two more features. First it is very fast in both training and testing documents. Second, it produces readable and understandable rules that can be easily modified by humans.

5. CONCLUSION AND FUTURE WORK

This paper approaches the problem of online text categorization using association rules. In particular, the study involves the application of FP-Growth algorithm to online news classification. The study provides evidence that association rule mining can be used for the construction of fast and effective classifiers for automatic text categorization. One major advantage of the association rule based classifier is that it does not assume that terms are independent and its training is relatively fast.
Furthermore, the rules are human understandable and easy to be maintained. Feature selection can be done by adding the weight of each term in the documents and pruning the terms with lower weight. The feature selection will reduce the number of terms as well as reduce the noisy of the terms. The feature selection techniques such as latent semantic analysis could improve the results.

**REFERENCES**


<table>
<thead>
<tr>
<th>category</th>
<th>ARC-FG</th>
<th>Bayes</th>
<th>Rocchio</th>
<th>CA .5</th>
<th>k-NN</th>
<th>SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
<td>5%</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>acq</td>
<td>8.4%</td>
<td>8.6%</td>
<td>7.1%</td>
<td>91.5%</td>
<td>92.1%</td>
<td>85.3%</td>
</tr>
<tr>
<td>corn</td>
<td>8.8%</td>
<td>8.6%</td>
<td>3.9%</td>
<td>47.3%</td>
<td>62.2%</td>
<td>87.7%</td>
</tr>
<tr>
<td>crude</td>
<td>2.1%</td>
<td>2.1%</td>
<td>1.4%</td>
<td>81.0%</td>
<td>81.5%</td>
<td>75.5%</td>
</tr>
<tr>
<td>earn</td>
<td>7.7%</td>
<td>7.0%</td>
<td>1.4%</td>
<td>95.3%</td>
<td>96.1%</td>
<td>96.1%</td>
</tr>
<tr>
<td>grain</td>
<td>0.6%</td>
<td>0.0%</td>
<td>9.9%</td>
<td>72.5%</td>
<td>79.5%</td>
<td>89.1%</td>
</tr>
<tr>
<td>interest</td>
<td>8.8%</td>
<td>8.8%</td>
<td>1.2%</td>
<td>58.0%</td>
<td>72.5%</td>
<td>49.7%</td>
</tr>
<tr>
<td>money-fx</td>
<td>0.1%</td>
<td>0.8%</td>
<td>0.8%</td>
<td>62.9%</td>
<td>67.6%</td>
<td>69.4%</td>
</tr>
<tr>
<td>ship</td>
<td>7.3%</td>
<td>7.7%</td>
<td>3.3%</td>
<td>78.5%</td>
<td>83.1%</td>
<td>80.2%</td>
</tr>
<tr>
<td>trade</td>
<td>9.0%</td>
<td>9.0%</td>
<td>4.8%</td>
<td>50.9%</td>
<td>77.4%</td>
<td>59.2%</td>
</tr>
<tr>
<td>wheat</td>
<td>4.9%</td>
<td>7.7%</td>
<td>5.6%</td>
<td>60.6%</td>
<td>79.4%</td>
<td>85.6%</td>
</tr>
<tr>
<td>micro-avg</td>
<td>8.6%</td>
<td>8.6%</td>
<td>6.5%</td>
<td>72.4%</td>
<td>79.9%</td>
<td>79.4%</td>
</tr>
<tr>
<td>macro-avg</td>
<td>0.8%</td>
<td>0.7%</td>
<td>0.6%</td>
<td>65.4%</td>
<td>79.1%</td>
<td>77.4%</td>
</tr>
</tbody>
</table>


[9] B. Liu, W. Hsu and Y. Ma, “Integrating classification and association rule mining”, In ACM Int. Conf. on knowledge Discovery and Data Mining.


Author’s Biography

Mrs. V. Srividhya received her M.Sc (Computer Science) from Madurai Kamaraj University in 1996. She obtained her M.Phil (Computer Science) from Bharathiar University in 2000. She is currently working towards the Ph.D degree with the specialization of Data mining. Since 2007, she has been associated with the Avinashilingam University for Women, Coimbatore, where she is currently an Asst. Professor in the Department of Computer Science. Her current research interests are in the area of Text mining and Web mining. She has presented and published number of papers in different International and National Conferences and journals.

Dr. R. Anitha obtained her Ph.D. from Periyar University, Salem. She is currently working as a Professor & Director, Department of MCA, K.S. Rangasamy college of Technology, Tiruchengode. Her area of Specialization is Digital Image Processing. She has presented and published number of papers in different International and National Conferences and journals.
A Novel Algorithm For Frequent Item-sets Mining Using Reduced Database Scan Approach

Mahesh H. Panchal 1 Bhagirath P. Prajapati 2

ABSTRACT

Association rule mining is one of the techniques of data mining by which valuable but hidden patterns (knowledge) are discovered from large amount of data. Mining of frequent item sets from which association rules are made, is a most challenging task. No. of algorithms had been developed for frequent item set mining, all differ in various aspects. In this paper a novel algorithm which we are calling as frequent 2-base is presented to mine frequent item sets in two database scans. The time taken by frequent 2-base is also compared with respect to various dimensions. Three dimensions are used to compare the time: database size, no. of items and average transaction size.

Keywords : Data Mining, Hidden Pattern, Minimum Support, Frequent Item Sets.

1. INTRODUCTION

In competitive market, any organization wants some interesting and hidden conclusions (results), trends, patterns from data stored in database. Data Mining is growing field which provides various methods by which those trends or patterns can be discovered. There are several methods available of data mining including association rule mining, classification, clustering etc. One can choose any one of them or any combination according to the requirement. Association Rule Mining (ARM) is one of the methods which generate closed associations among data items of database. If ARM is applied on market basket data, discovered associations among data items indicates that during one visit of market if customer purchases some items along with them which other items he may purchase.

Mining association rules is a two step process\cite{1}\cite{2}. (1) Generating all items which are frequently purchased by customers in market basket problem. The item sets which occur more than user defined threshold value called min_sup are called frequent item sets. (2) Generating all strong association among frequent item sets. The association is called strong if association satisfies another user defined threshold value called min_conf. The problem of finding frequent item sets attracted the attention of many researchers. As a result, at present there exist several algorithms for the same including the ones based on variations of apriori, based on depth first search approach, tree based algorithms\cite{3}\cite{4}\cite{5}.

When any algorithm is finding the frequent item sets, the database remains into secondary storage. Of course, the database is very large so, one of the parameter one should take into consideration in any algorithm is how many i/o access are performed i.e. how many times the database is scanned\cite{14}\cite{15}. The apriori algorithm scans the whole database \(k+l\) times, where \(k\) is the length of longest frequent item set. Some other algorithms are able to find same set of frequent patterns in two scans of database only irrespective of database size and length of frequent item sets.

1 Asst. Prof., Dept. of Comp Engg., Kalol Institute of Tech & Research Centre, Gujarat. Email : mkhpanchal@gmail.com
2 Lecturer, Dept. of Comp. Engg., A.D. Patel Institute of Technology, Gujarat. Email : bhagirath_123@yahoo.co.in
In this paper an algorithm called frequent 2-base is shown which requires two scans of database and finds set of frequent patterns in less amount of time compared to other algorithms which also requires same no. of scans. In frequent 2-base first all frequent 2-item sets are derived by scanning the whole database once. From frequent 2-item sets all possible higher order candidate item sets are found. Then from set of candidate item sets of all orders, frequent item sets are taken out by scanning the database second time. It had been verified that no frequent item set of any order is missed out by the algorithm frequent 2-base.

The time taken by frequent 2-base algorithm is compared by various parameters namely size of database (D), no. of data items (N) and average length of transaction (AT). All the statistical results are presented in tabular and graphical form.

2. ASSOCIATION RULE

Discovering association rules is at the heart of data mining. It detects hidden linkages of otherwise seemingly unrelated data. These linkages are rules. Those that exceed a certain threshold are deemed interesting. Interesting rules allow actions to be taken based upon data pattern. They can also help making and justifying decisions. In association rule mining there are two measurements, support and confidence. The support corresponds to the frequency of the pattern while confidence indicates rule’s strength. A typical example of an association rule created by data mining often termed to as “market basket data” is: “80% of customers who purchase bread also purchase butter.” An association rule is defined as: 

Let D is set of all distinct data items available in database. T is set of all transactions \( t_1, t_2, t_3, \ldots, t_n \). Every transactions \( t_i \), where \( i \) may be \( 1,2,3,\ldots,n \) contains some data items from D. The association rule is of form \( X \rightarrow Y \) where X and Y is subset of D and \( X \cap Y = \emptyset \). It is said that association rule \( X \rightarrow Y \) holds in database if and only if support of \( X \rightarrow Y \) is above min_sup and confidence is above min_conf.

Support of \( X \rightarrow Y \) is percentage of transactions out of total transactions, which contain all the data items available in item-sets X and Y. Confidence of \( X \rightarrow Y \) is percentage of transactions which contain all the items in Y out of the transactions which contain all the items in X. Suppose bread \( \rightarrow \) milk holds in some market basket database with total 100 transactions with support 60% and confidence 75%. It is interpreted that 60% of transactions out of 100 contain both bread and milk and if bread is purchased than in 75% of those cases milk is also purchased.

Given a user defined minimum support and minimum confidence, the problem of mining association rules is to find all the associations rules whose support and confidence are larger than the minimum support (min_sup) and minimum confidence (min_conf). Thus, the approach can be broken into two sub-problems as follows:

1. Finding the frequent item sets which have support above the predetermined minimum support.
2. Deriving strong association rules, based on each frequent item set, which have confidence more than the minimum confidence.

2.1 Frequent Item-Set Mining

The task of discovering all frequent item sets is quite challenging. The search space is exponential in the number of items occurring in the database. The support threshold limits the output to a hopefully reasonable subspace. Also, such databases could be massive, containing millions of transactions, making support counting a tough problem.
Apriori is basic algorithm for frequent item sets mining. In first step it finds \( L_1 \), i.e. set of frequent 1-item sets by scanning the database first time. From \( L_1 \) it finds set of candidate 2-item sets \( C_2 \). Again by scanning the database it finds \( L_2 \) from \( C_2 \). This process is repeated until no more \( C_k \) is possible from \( L_{k-1} \) or no further \( L_k \) is found from \( C_k \). Apriori uses the prior knowledge by pruning out to generate those candidate item sets whose at least one subset is not frequent. The main drawback of this algorithm is, it is scanning the database \( k+1 \) times where \( k \) is the length of longest frequent item set. As a result of it i/o cost is increased as the length of longest frequent item set is increased. The most outstanding improvement over Apriori would be a method called FP-growth (frequent pattern growth) that succeeded in eliminating candidate generation\[^8\][^9\]. It adopts a divide and conquer strategy by (1) compressing the database representing frequent items into a structure called FP-tree (frequent pattern tree) that retains all the essential information and (2) dividing the compressed database into a set of conditional databases, each associated with one frequent item set and mining each one separately. It scans the database only twice. In the first scan, all the frequent items and their support counts (frequencies) are derived and they are sorted in the order of descending support count in each transaction. In the second scan, items in each transaction are merged into a prefix tree and items (nodes) that appear in common in different transactions are counted. Each node is associated with an item and its count. Nodes with the same label are linked by a pointer called node-link. Since items are sorted in the descending order of frequency, nodes closer to the root of the prefix tree are shared by more transactions, thus resulting in a very compact representation that stores all the necessary information. Pattern growth algorithm works on FP-tree by choosing an item in the order of increasing frequency and extracting frequent item sets that contain the chosen item by recursively calling itself on the conditional FP-tree. FP-growth is an order of magnitude faster than the original Apriori algorithm.

3. FREQUENT 2-BASE ALGORITHM

This section represents the algorithm frequent 2-base. The complete algorithm is decomposed into following phases: (1) generation of candidate 2-item sets. (2) finding frequent 2-item sets. (3) generating all possible higher order candidate item sets. (4) generation of all higher order frequent item sets. Out of these four phases, first and third require database scan. The output generated by each phase is given as input to next phase. Input of first phase is original database and output of last phase is frequent item sets.

3.1 Generation Of Candidate 2-item Sets

Section 3.1.1 shows pseudo code to generate candidate 2-item sets from given market basket database. That code is explained in section 3.1.2.

3.1.1 Pseudo Code

1. \( \text{temp} = \text{null} \).
2. Repeat for \( i=1 \) to \( n \)
3. Repeat for each candidate 2-item set \( X_2 \)
4. If \( X_2 \) is not in \( \text{temp} \)
5. \( X_2.\text{count} = 1 \)
6. \( \text{temp} = \text{temp} \cup X_2 \)
7. If \( X_2 \) is in \( \text{temp} \)
8. \( X_2.\text{count} = X_2.\text{count} + 1 \)
9. end.
10. end.
11. copy all candidate 2-item sets from \( \text{temp} \) to \( C_2 \).
12. \( \text{temp} = \text{null} \).

3.1.2 Explanation

This step of algorithm is to generate candidate 2-item sets from whole database. In above pseudo code, \( n \) is no. of transactions; \( \text{temp} \) is temporary buffer which contains
candidate 2-item sets before finally they will be put in $C_2$. Whole database is first time scanned to find candidate 2-item sets. Suppose some transaction contains the items \{B,C,D,F\} then its candidate 2-item sets are \{B,C\}, \{B,D\}, \{B,F\}, \{C,D\}, \{C,F\} and \{D,F\}. For each candidate 2-item set a separate count is maintained which contains the frequency of that item-set in database.

While implementing this phase, a linked list is maintained which consists of a node for each candidate 2-item set. Support count for each candidate 2-item set is calculated and stored in a field in each node. When a new candidate item-set is found a new node is inserted at the end of linked list otherwise if a node for some item-set is already there then its support count is incremented by one.

3.2 Finding Frequent 2-item Sets
Section 3.2.1 shows pseudo code to find frequent 2-item sets from candidate 2-item sets found from previous phase. That code is explained in section 3.2.2.

3.2.1 Pseudo Code
1. Repeat for each candidate 2-item set $X_2$
2. If ($X_2$.count < min_sup )
3. $C_2 = C_2 - X_2$
4. end.

3.2.2 Explanation
The candidate 2-item sets whose frequency is less than minimum requirement are removed in this phase. The set of remaining candidate 2-item sets are called frequent 2-item sets.

3.3 Generating All Possible Higher Order Candidate Item Sets
Section 3.3.1 shows pseudo code to generate higher order candidate item sets from frequent 2-item sets. That code is explained in section 3.3.2.

3.3.1 Pseudo Code
1. Repeat while ($C_k$ is not empty and $k>=2$)
2. $C_{k+1} = C_k * C_k$
3. increment $k$.
4. end.

3.3.2 Explanation
From previous section 3.2, no. of frequent 2-item sets is found. If conventional method like apriori is used then first frequent 2-item sets should be found from candidate 2-item sets, then from frequent 2-item sets, candidate 3-item sets are found. In general from candidate k-item sets first frequent k-item sets are found by date base scan and from frequent k-item sets, candidate (k+1) item sets are found. It is obvious that if maximum k frequent item set is possible then (k+1) database scans are to be performed. This is time consuming task.

In order to save the execution time of an algorithm, the number of database scans should be reduced with out any harmful effect on results. This can be done with scan reduction technique. From frequent 2-item sets, first candidate 3-item sets are found based on property of prior knowledge. It means for any item set to be frequent, all of its subsets must be frequent. It follows that if for some 3-item set to be frequent, all of its 2-item set subsets must be frequent. By this way all the candidate 3-item sets are found. They can not be called as frequent 3-item sets because they may not satisfy required support threshold. It will be checked with second database scan. Then from candidate 3-item sets, candidate 4-item sets are found and so on. At any time an item set $X_k$ is put in candidate k-item set only if all its subsets are there in candidate (k-1) item sets. This process is repeated for all higher order item sets until no higher order item set is possible to generate.
In our implementation of this phase, first all candidate 3-item sets are found from frequent 2-item sets. For each one of them a new node is inserted in existing linked list of frequent 2-item sets at the end. Then candidate 4-item sets, 5-item sets etc. are found and inserted in same linked list.

3.4 Finding All Higher Order Frequent Item Sets

Section 3.4.1 shows pseudo code to find all order frequent item sets from respective candidate item sets found from previous phase. That code is explained in section 3.4.2.

3.4.1 Pseudo Code

1. Repeat for i=1 to n
2. Find all subsets having length at least 3 of item-set present in transaction i.
3. For each subset X_k in C_k
4. X_k.count = X_k.count + 1
5. end.
6. end.
7. Repeat for k=3 to m
8. Repeat for each candidate item set X_k
9. If (X_k.count < min_sup )
10. C_k = C_k – X_k
11. end.
12. end.

3.4.2 Explanation

This step of algorithm is to prune out infrequent item sets from any order. In this code, n is no. of transactions, m is no. of nodes in linked list. Lines 1 to 6 finds all subsets of length at least 3 for item set available in each transaction. Then support count for each subset is incremented by one in respective node. Lines 7 to 12 compare the support count for each node with minimum support required and delete the nodes whose support is less. So, after this phase, frequent item sets of all orders are present with support count in linked list.

4. Statistical Analysis

The frequent 2-base algorithm is implemented in Microsoft visual C++ 6.0. The code is executed for different size of database, different no. of items and different average transaction length. Each time min_sup is kept constant of 20. The required database is generated with pure randomness by synthetic database generator available in ARTool. All testing is done on single system which has intel core 2 duo 2.53 GHz processor, 1GB RAM and 80GB SATA HDD. In section 4.1, a comparative table is shown and briefly explained. Section 4.2 shows graphical presentation of execution time by comparing two dimensions each time and keeping value of third dimension constant.

4.1 Experimental Results

The execution time of frequent 2-base algorithm is calculated for different dimensions (parameters) like size of database (D), no. of data items (N) and average length of transaction (AT). Table (1) shows the overall results. Three different values are used for each dimension in testing. For database size D values are 100, 1000 and 10000. D=1000 means in database there are 1000 transactions available. For no. of items N values are 10, 15 and 20. N=15 means in database there are 15 different items. For average length of transaction AT values are 3, 5 and 7. All execution times are in millisecond. The values in brackets show the total no. of frequent item sets. For example, when D=10000 and N=15, if AT=3 then 27 frequent item sets are mined in 606ms, if AT=5 then 192 frequent item sets are mined in 5197ms and if AT=7 then 265 frequent item sets are mined in 23359ms.

4.2 Graphical Comparison

The execution time of frequent 2-base is compared graphically for three different values of two dimensions at a time by keeping value of third dimension constant.
Fig. 1 shows the comparison of database size and no. of items. In (a), (b) and (c) of Fig. 1 average length of transaction is kept constant of 3, 5 and 7 respectively.

Fig. 2 shows the comparison of database size and average length of transaction. In (a), (b) and (c) of Fig. 2 no. of items is kept constant of 10, 15 and 20 respectively.

Fig. 3 shows the comparison of no. of items and average length of transaction. In (a), (b) and (c) of Fig. 3 database size is kept constant of 100, 1000 and 10000 respectively.

5. CONCLUSION

An algorithm frequent 2-base finds all frequent item sets in two database scans. Here linked list is used as a data structure for maintaining frequent item sets. The main attraction of this algorithm is, it can be effectively used for mining frequent item sets from seasonal market basket database where some transactions or part of them are repeated over time. The general trend is that when size of database and no. of items increases, the execution time naturally increases. But it may not be the case all time. It also depends on third and most interesting parameter called average length of transaction. One can easily analyze the code of the algorithm by varying the values of all these parameters.
A Novel Algorithm For Frequent Item-sets Mining Using Reduced Database Scan Approach

Figure 2: Database Size Vs. Average Length of Transaction

(a) D = 100
(b) N = 15
(c) N = 20

Figure 3: No. of items Vs. Average Length of Transaction

(a) D = 1000
(b) D = 10000
(c) D = 100000

Table 1: Execution Time (In Millisecond) For Various Values Of Dimensions

<table>
<thead>
<tr>
<th>Data base size (D)</th>
<th>No. of Items (N)</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>49, (49)</td>
<td>16(50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49, (249)</td>
<td>47(72)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 (146)</td>
<td>62(181)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 (146)</td>
<td>256(281)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>406(41)</td>
<td>3186(1013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>289(145)</td>
<td>78(41)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>441(285)</td>
<td>820(208)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10000</td>
<td>1268(120)</td>
<td>731(26)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>506(32)</td>
<td>731(26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>820(336)</td>
<td>34727(627)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES


[9] J. Han, J. Pei and Y. Yin, “Mining Frequent Patterns without Candidate Generation”, Proc. of 2000 ACM-SIGMOD Int. Conf. on Management of Data, May 2000.


Author’s Biography

Prof. Mahesh H. Panchal is Head and Assistant Professor with Department of Computer Engineering, Kalol Institute of Technology & Research Centre, Gujarat. He did his graduation from Gujarat University, 2002 and
A Novel Algorithm For Frequent Item-sets Mining Using Reduced Database Scan Approach

post graduation from Sardar Patel University, 2009 in computer engineering. He has a teaching experience of around 8 years. He had published a research paper in national journal, two research papers in proceeding and a paper presented in national conference. His area of interest include Database Management System, Data Mining & Warehousing, Real Time Systems.

Prof. Bhagirath P. Prajapati is Assistant Professor with Department of Computer Engineering, A. D. Patel Institute of Technology, Gujarat. He did his graduation and post graduation from Sardar Patel University 2003, 2010 respectively in computer engineering. He has a teaching experience of around 6 years. He had published a research paper in proceeding and a paper presented in national conference. His area of interest include Operating System, Windows Programming and Database Management System.
A Novel Approach for Content Based Image Retrieval

Srinivasa Kumar Devireddy

ABSTRACT

The importance of an effective technique in searching and retrieving images from the huge collection cannot be overemphasized. One approach for indexing and retrieving image data is using manual text annotations. The annotations can then be used to search images indirectly. But there are several problems with this approach. First, it is very difficult to describe the contents of an image using only a few keywords. Second, the manual annotation process is very subjective, ambiguous, and incomplete. Those problems have created great demands for automatic and effective techniques for content-based image retrieval (CBIR) systems. Most CBIR systems use low-level image features such as color, texture, shape, edge, etc., for image indexing and retrieval. It’s because the low-level features can be computed automatically. Content Based Image Retrieval has emerged during the last several years as a powerful tool to efficiently retrieve images visually similar to a query image. The main idea is to represent each image as a feature vector and to measure the similarity between images with distance between their corresponding feature vectors according to some metric. Finding the correct features to represent images with, as well as the similarity metric that groups visually similar images together, are important steps in the construction of any CBIR system.

Keywords: CBIR, Image, Feature, Indexing, Matching, GIS, Quantization, Texture.

1. INTRODUCTION

The field of image retrieval has been an active research area for several decades and has been paid more and more attention in recent years as a result of the dramatic and fast increase in the volume of digital images. The development of Internet not only cause an explosively growing volume of digital images, but also give people more ways to get those images.

There were two approaches to content-based image retrieval initially. The first one is based on attribute representation proposed by database researchers where image contents are defined as a set of attributes which are extracted manually and are maintained within the framework of conventional database management systems [2][3]. Queries are specified using these attributes. This obviously involves high-level of image abstraction. The second approach which was presented by image interpretation researchers depends on an integrated feature-extraction / object-recognition subsystem to overcome the limitations of attribute-based retrieval. This system automates the feature-extraction and object-recognition tasks that occur when an image is inserted into the database [15][18]. This automated approaches to object recognition are computationally expensive, difficult and tend to be domain specific. There are two major categories of features. One is basic which is concerned with extracting boundaries of the image and the other one is logical which defines the image at various levels of details. Regardless of which approach is used, the retrieval in content-based image retrieval is done by color, texture, sketch, shape, volume, spatial constraints,
browsing, objective attributes, subjective attributes, motion, text and domain concepts.

1.1 The Growth of Digital Imaging

The use of images in human communication is hardly new – our cave-dwelling ancestors painted pictures on the walls of their caves, and the use of maps and building plans to convey information almost certainly dates back to pre-Roman times. But the twentieth century has witnessed unparalleled growth in the number, availability and importance of images in all walks of life. Images now play a crucial role in fields as diverse as medicine, journalism, advertising, design, education and entertainment. Technology, in the form of inventions such as photography and television, has played a major role in facilitating the capture and communication of image data. But the real engine of the imaging revolution has been the computer, bringing with it a range of techniques for digital image capture, processing, storage and transmission which would surely have started even pioneers like John Logie Baird. The involvement of computers in imaging can be dated back to 1965, with Ivan Sutherland’s Sketchpad project, which demonstrated the feasibility of computerised creation, manipulation and storage of images, though the high cost of hardware limited their use until the mid-1980s. Once computerised imaging became affordable, it soon penetrated into areas traditionally depending heavily on images for communication, such as engineering, architecture and medicine. Photograph libraries, art galleries and museums, too, began to see the advantages of making their collections available in electronic form[7]. The creation of the World-Wide Web in the early 1990s, enabling users to access data in a variety of media from anywhere on the planet, has provided a further massive stimulus to the exploitation of digital images. The number of images available on the Web was recently estimated to be between 10 and 30 million a figure which some observers consider to be a significant underestimate.

1.2 CBIR

Content Based Image Retrieval is the retrieval of images based on visual features such as colour and texture.

Reasons for its development are that in many large image databases, traditional methods of image indexing have proven to be insufficient, laborious, and extremely time consuming[15][18]. These old methods of image indexing, ranging from storing an image in the database and associating it with a keyword or number, to associating it with a categorized description, have become obsolete. This is not in CBIR [5][6]. In CBIR, each image that is stored in the database has its features extracted and compared to the features of the query image. It involves two steps:

* Feature Extraction: The first step in the process is extracting image features to a distinguishable extent.
* Matching: The second step involves matching these features to yield a result that is visually similar.

1.3 Existing System

Early techniques were not generally based on visual features but on the textual annotation of images. In other words, images were first annotated with text and then searched using a text-based approach from traditional database management systems. Text-based image retrieval uses traditional database techniques to manage images [5][6][7]. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. However, since automatically generating descriptive texts for a wide spectrum of images is not feasible, most text-based image retrieval systems require manual annotation of images. Obviously, annotating images manually is a cumbersome and expensive task.
for large image databases, and is often subjective, context-sensitive and incomplete. As a result, it is difficult for the traditional text-based methods to support a variety of task-dependent queries.

1.4 Proposed System

The solution proposed is to extract the primitive features of a query image and compare them to those of database images. The image features under consideration were colour, texture and shape. Thus, using matching and comparison algorithms, the colour, texture and shape features of one image are compared and matched to the corresponding features of another image. This comparison is performed using colour, texture and shape distance metrics. In the end, these metrics are performed one after another, so as to retrieve database images that are similar to the query. The similarity between features is to be calculated using Distance algorithm.

2. APPLICATIONS OF CBIR

A wide range of possible applications for CBIR technology has been identified. Potentially fruitful areas include: Crime prevention, the military, Fashion and interior design, Journalism and advertising, Medical diagnosis, Geographical information and remote sensing systems, Web searching.

Crime Prevention: Law enforcement agencies typically maintain large archives of visual evidence, including past suspects’ facial photographs (generally known as mug shots), fingerprints, tyre treads and shoeprints. Whenever a serious crime is committed, they can compare evidence from the scene of the crime for its similarity to records in their archives. Strictly speaking, this is an example of identity rather than similarity matching, though since all such images vary naturally over time, the distinction is of little practical significance. Of more relevance is the distinction between systems designed for verifying the identity of a known individual (requiring matching against only a single stored record), and those capable of searching an entire database to find the closest matching records [9].

The Military: Military applications of imaging technology are probably the best-developed, though least publicized. Recognition of enemy aircraft from radar screens, identification of targets from satellite photographs, and provision of guidance systems for cruise missiles are known examples – though these almost certainly represent only the tip of the iceberg. Many of the surveillance techniques used in crime prevention could also be relevant to the military field.

Fashion and Interior Design: Similarities can also be observed in the design process in other fields, including fashion and interior design. Here the designer has to work within externally-imposed constraints, such as choice of materials. The ability to search a collection of fabrics to find a particular combination of colour or texture is increasingly being recognized as a useful aid to the design process.

Journalism and Advertising: Both newspapers and stock shot agencies maintain archives of still photographs to illustrate articles or advertising copy. These archives can often be extremely large (running into millions of images), and dauntingly expensive to maintain if detailed keyword indexing is provided. Broadcasting corporations are faced with an even bigger problem, having to deal with millions of hours of archive video footage, which are almost impossible to annotate without some degree of automatic assistance.

Medical Diagnosis: The increasing reliance of modern medicine on diagnostic techniques such as radiology, histopathology, and computerised tomography has resulted in an explosion in the number and importance of
medical images now stored by most hospitals. While the prime requirement for medical imaging systems is to be able to display images relating to a named patient, there is increasing interest in the use of CBIR techniques to aid diagnosis by identifying similar past cases.

Geographical Information Systems (GIS) and Remote Sensing: Although not strictly a case of image retrieval, managers responsible for planning marketing and distribution in large corporations need to be able to search by spatial attribute (e.g. to find the 10 retail outlets closest to a given warehouse). And the military are not the only group interested in analysing satellite images. Agriculturalists and physical geographers use such images extensively, both in research and for more practical purposes, such as identifying areas where crops are diseased or lacking in nutrients – or alerting governments to farmers growing crops on land they have been paid to leave lying fallow.

Web Searching: Cutting across many of the above application areas is the need for effective location of both text and images on the Web, which has developed over the last five years into an indispensable source of both information and entertainment. Text-based search engines have grown rapidly in usage as the Web has expanded; the well-publicized difficulty of locating images on the Web indicates that there is a clear need for image search tools of similar power. Paradoxically, there is also a need for software to prevent access to images which are deemed pornographic.

3. Literature Survey

3.1 Image and Digital Imaging

It is said that one image is worth a thousand words. Visual information accounts for about 90% of the total information content that a person acquires from the environment through his sensory systems. This reflects the fact that human being relies heavily on his highly developed visual system compared with other sensory pathways. The external optical signal is perceived by eyes, and then converted into neural signal; the corresponding neural subsystem specialized for visual system is specially organized to detect subtle image features and perform high-level processing, which is further processed to generate object entities and concepts. The anatomy of the visual system explains from the structure aspect why visual information is so important to human cognition. The cognitive functions that such a system must support include the capability to distinguish among objects, their positions in space, motion, sizes, shapes, and surface texture [6]. Some of these primitives can be used as descriptors of image content in machine vision research.

3.1.1 Taxonomy of Images

There are two formats, in which visual information can be recorded and presented – static image, and motion picture, or video. Image is the major focus of research interest in digital image processing and image understanding. Although a relatively recent development, computerized digital image processing has attracted much attention and shed lights to a broad range of existing and potential applications. This is directly caused by rapid accumulation of image data, a consequence of exponential increases of digital storage capacity and computer processing power. There are several major types of digital images depending on the elemental constituents that convey the image content. Images can take the form of:

1. Printed text and manuscript. Some examples of the kind are micro-films of old text documents, photograph of handwriting.
2. Line sketch, including diagrams, simple line graphs
3. Halftones. Images are represented by a grid of dots of variable sizes and shapes.

5. Mixture of above.

Among all above, continuous tone or photographic images are most common in the digital imaging practice and of major concern of content-based image retrieval. They used to be generated by converting from other media using a scanner. The process is labour intensive and costly. It was estimated that the image capturing and the subsequent manual indexing may account for 90 percent of the total cost of building an image database. This was the situation a few years ago. Now, digital cameras are becoming very popular and images are also being converted from analog electronic format to digital format.

At the conceptual level, an image is a representation of its target object(s). According to the Webster’s 3rd New International Dictionary, an image is “the optical counterpart of an object ... a mental picture, a mental conception ...” The definition reflects the fact that an image has its content, which captures the optical or mental properties of an object; with its format varying across different kinds of media. It is determined by how the optical properties are quantized, or the degree of mental abstractions that is required. Here are some examples of image according to the above broader definition:

1. An image can be an array of pixel values stored in uncompressed bitmap digital image format. In this format, each value represents the color intensity at discrete points or pixels. A well-known example of this is Microsoft’s BMP format. Although BMP format allows for pixel packing and run-length encoding to achieve certain level of compression, its uncompressed version is more popular.

2. Popular Internet standard image formats see more extensive image transformation and compression, such as GIF and JPEG. The GIF standard defines a color degeneration process, which maps the colors in an image into no more than 256 new colors.

3. Specially defined signature file stores specifically extract images features in numeric or Boolean format, indicating the presence (or non-presence) and strength of the features. This is a very compact representation of image that is targeted for fast retrieval instead of display, archival, etc. Only the essential image features are conserved in the signature file and they are algorithm dependent. This allows for easy indexing and fast search for matching features of the query. An example of this can be found in image coding method using vector quantization (VQ), in which image blocks are coded according to a carefully chosen codebook. If the image blocks are similar to each other or the images in a set bear significant similarity, higher compression ratio can usually be achieved than the general-purpose compression algorithms such as GIF, JPEG.

4. Textual annotation can also be thought of as an instantiation of mental image, and sometimes, the descriptors can be coded by a predefined convention, or a thesaurus. The fact that two visually different images can convey the same concept and different concepts may present in images that share many similar optimal properties brings about a gap between image retrieval by content, and retrieval by concept.

3.2 Visual Features

Feature extraction plays an important role in content-based image retrieval to support for efficient and fast retrieval of similar images from image databases. Significant features must first be extracted from image data. Retrieving images by their content, as opposed to external features, has become an important operation. A fundamental ingredient for content based image retrieval is the technique used for comparing images. There are
two general methods for image Comparison: intensity based (color and texture) and geometry based (shape).
So we will concentrate on these features.

3.2.1 Colour

One of the most important features that make possible the recognition of images by humans is colour. Colour is a property that depends on the reflection of light to the eye and the processing of that information in the brain. We use colour everyday to tell the difference between objects, places, and the time of day. Usually colours are defined in three dimensional colour spaces. These could either be RGB (Red, Green, and Blue), HSV (Hue, Saturation, and Value) or HSB (Hue, Saturation, and Brightness). The last two are dependent on the human perception of hue, saturation, and brightness. Most image formats such as JPEG, BMP, GIF, use the RGB colour space to store information. The RGB colour space is defined as a unit cube with red, green, and blue axes. Thus, a vector with three co-ordinates represents the colour in this space. When all three coordinates are set to zero the colour perceived is black. When all three coordinates are set to 1 the colour perceived is white. The other colour spaces operate in a similar fashion but with a different perception.

3.2.2 Transformation and Quantization

The color regions are perceptually distinguishable to some extent. The human eye cannot detect small color different and may perceive these very similar colors as the same color. This leads to the quantization of color, which means that some pre-described colors will be present on the image and each color is mapped to some of these pre-described colors. One obvious consequence of this is that each color space may require different levels of quantized colors, which is nothing but a different quantization scheme. In below, the effect of color quantization is illustrated. Figure (a) is the original image with RGB color space and (b) is the image produced after transformation into HSV color space and quantization.

![Figure 1: Transformation, Quantization of Tiger Image](image)

3.2.3 Methods of Representation

The main method of representing colour information of images in CBIR systems is through colour histograms. A colour histogram is a type of bar graph, where each bar represents a particular colour of the colour space being used. We can get a colour histogram of an image in the RGB or HSV colour space. The bars in a colour histogram are referred to as bins and they represent the x-axis. The number of bins depends on the number of colours there are in each bin. In other words how many pixels in an image are of a particular colour.

3.2.4 Color Content Extraction

One of the widely used methods for querying and retrieval by color content is color histograms. The color histograms are used to represent the color distribution in an image mainly, the color histogram approach counts the number of occurrences of each unique color on a sample image. Since an image is composed of pixels and each pixel has a color, the color histogram of an image can be computed easily by visiting every pixel once. Smith and Chang proposed colorsets as an opponent to color histograms. The colorsets are binary masks on color histograms and
they store the presence of colors as 1 without considering their amounts. For the absent colors, the colorsets store 0 in the corresponding bins. The colorsets reduce the computational complexity of the distance between two images. Besides, by employing colorsets region-based color queries are possible to some extent. On the other hand, processing regions with more than two or three colors is quite complex. Another image content storage and indexing mechanism is color correlograms. It involves an easy-to-compute method and includes not only the spatial correlation of color regions but also the global distribution of local spatial correlation of colors. In fact, a color correlogram is a table each row of which is for a specific color pair of an image. The $k$-th entry in a row for color pair $(i; j)$ is the probability of finding a pixel of color $j$ at a distance $k$ from a pixel of color $i$. The method resolves the drawbacks of the pure local and pure global color indexing methods since it includes local spatial color information as well as the global distribution of color information.

3.2.5 Texture

Texture is that innate property of all surfaces that describes visual patterns, each having properties of homogeneity. It contains important information about the structural arrangement of the surface, such as; clouds, leaves, bricks, fabric, etc. It also describes the relationship of the surface to the surrounding environment. In short, it is a feature that describes the distinctive physical composition of a surface. Texture properties include: (i) Coarseness, (ii) Contrast, (iii) Directionality, (iv) Line-likeness, (v) Regularity & (v) Roughness.

Texture is one of the most important defining features of an image. It is characterized by the spatial distribution of gray levels in a neighbourhood [6][8][13]. In order to capture the spatial dependence of gray-level values, which contribute to the perception of texture, a two-dimensional dependence texture analysis matrix is taken into consideration. This two-dimensional matrix is obtained by decoding the image file; jpeg, bmp, etc.

Figure 2 : Examples of Textures

3.3 MPEG-7 Texture Descriptors

The MPEG-7 multimedia content description interface involves three texture descriptors for representing texture regions in images [1], namely

- **Texture browsing descriptor** to characterize perceptual directionality, regularity, and coarseness of a texture,
- **Homogeneous texture descriptor** (HTD) to quantitatively characterize homogeneous texture regions for similarity retrieval using local spatial statistics of the texture obtained by scale- and orientation-selective Gabor filtering, and
- **Edge histogram descriptor** to characterize non-homogeneous texture regions.

3.4 Distance Measure Techniques

3.4.1 Histogram Intersection Method

In the Histogram Intersection technique, two normalized histograms are intersected as a whole, as the name of the technique implies. The similarity between the histograms is a floating point number between 0 and 1. Equivalence is designated with similarity value 1 and the similarity between two histograms decreases when the similarity
value approaches to 0. Both of the histograms must be of the same size to have a valid similarity value. Let \( H_1[1:n] \) and \( H_2[1:n] \) denote two histograms of size \( n \), and \( SH_1: H_2 \) denote the similarity value between \( H_1 \) and \( H_2 \). Then, \( SH_1: H_2 \) can be expressed by the distance between the histograms \( H_1 \) and \( H_2 \) as:

\[
SH_1: H_2 = \frac{\sum_{i=1}^{n} \min(H_1[i], H_2[i])}{\max(|H_1|, |H_2|)}
\]

In the system, this technique is employed for similarity calculations as a result of texture vector and color histogram comparisons between database images and query image.

4. General Schema of Content Based Image Retrieval

The block diagram consists of following main blocks - digitizer, feature extraction, image database, feature database, and matching and multidimensional indexing.

Function of each block is as follows.

**Digitizer:** To add new images in image database or query images which are acquired from CCD Camera, X-ray imaging system, microdensitometers, image dissectors, vision cameras etc. are needed to be digitized, so that computer can process those images.

**Image Database:** The Comparison between Query image and images from image database can be done directly pixel by pixel which will give precise match but on the other hand, recognizing objects entirely at query time will limit the retrieval speed of the system, due to the high expense of such computing. Generally this crude method of comparison is not used, but image database, which contains raw images, is required for visual display purpose.

**Feature Extraction:** To avoid above problem of pixel-by-pixel comparison next abstraction level for representing images is the feature level. Every image is characterized by a set of features such as Texture, Color, Shape and others. Extract these features at the time of injecting new image in image database. Then summarize these features in a reduced set of \( k \) indexes and store it in Image feature database. The query image is processed in the same way as images in the database. Matching is carried out on the feature database.

**Image Matching and Multidimensional Indexing:**

Extracted features of query image are compared with features, which are stored in image feature database. To achieve fast retrieval speed and make the retrieval system truly scalable to large size image collections and effective multidimensional indexing is indispensable part of the whole system. The system selects the \( N \) images having the greatest overall similarities to the query image.

5. Theoretical Analysis

The basic idea behind content-based image retrieval is that, when building an image database, or retrieving an image from the database, we first extract feature vectors from images (the features can be color, shape, and texture), then store the vectors in another database for future use [2][3]. When given a query image, we similarly extract its feature vectors, and match these vectors with
those already in the database, if the distance between two images feature vectors is small enough; we consider the corresponding image in the database match the query. The search is usually based on similarity rather than on exact match, and the retrieval results are then ranked according to a similarity index. Usually, a group of similar target images are presented to users.

Recent efforts in image retrieval applications have focused on indexing a few specific visual dimensions of the images, such as color, texture, shape, motion and spatial information. However without integrating these visual dimensions, the current content-based techniques have limited capacity to satisfactorily retrieve images. Many difficult problems in image retrieval systems remain to be investigated. The explosive proliferation of “unconstrained” digital imagery in the form of images, graphics, and videos, due to the improved accessibility of computer technology and the main-stream acceptance of the world-Wide Web as a variable medium for publishing, advertising and communicating has created an immediate need for efficient and effective tools for cataloguing, indexing, managing, compressing and searching for the “unconstrained” visual information. A significant gap exists between the ability of computers to analyze images and videos at the feature level (colors, textures, shapes) compared to the inability at the semantic-level (objects, scenes, people, moods, artistic value). Closing this gap by improving technologies for image understanding would certainly improve the image search and retrieval systems. However, the solution of this “large” problem remains distant while much of the current technology consists of solutions to “small” problems (for example, image segmentation, color constancy, shape from texture, video shot detection, and so forth). However there are new opportunities to climb the semantic ladder in today’s divers media enjoinments.

The final difficulty limiting progress in image retrieval concerns system evaluation. Unless there are reliable and widely accepted ways of measuring effectiveness of new technique, it will be impossible to judge whether they represent any advancement on existing methods. These will inventible limit the progress. For an image retrieval system to be successful for geographical images, we need to develop approaches robust in spatial pattern characterization, rotation invariant, and taking into account of the scale effect. This project can be thought of having two phases of operations [9][10]. The first one consists of different types of image storage techniques, the second one is having the retrieving the image from existing database.

What kinds of query are users likely to put to an image database? To answer this question in depth requires a detailed knowledge of user needs – why users seek images, what use they make of them, and how they judge the utility of the images they retrieve. As we show in below, not enough research has yet been reported to answer these questions with any certainty. Common sense evidence suggests that still images are required for a variety of reasons, including:

- illustration of text articles, conveying information or emotions difficult to describe in words,
- display of detailed data (such as radiology images) for analysis,
- formal recording of design data (such as architectural plans) for later use.

Access to a desired image from a repository might thus involve a search for images depicting specific types of object or scene, evoking a particular mood, or simply containing a specific texture or pattern. Potentially, images have many types of attribute which could be used for retrieval, including:
• the presence of a particular combination of colour, texture or shape features (e.g. green stars);
• the presence or arrangement of specific types of object (e.g. chairs around a table);
• the depiction of a particular type of event (e.g. a football match);
• the presence of named individuals, locations, or events (e.g. the Queen greeting a crowd);
• subjective emotions one might associate with the image (e.g. happiness);
• metadata such as who created the image, where and when.

Level 1 comprises retrieval by primitive features such as colour, texture, shape or the spatial location of image elements. Examples of such queries might include “find pictures with long thin dark objects in the top left-hand corner”, “find images containing yellow stars arranged in a ring” – or most commonly “find me more pictures that look like this”. This level of retrieval uses features which are both objective, and directly derivable from the images themselves, without the need to refer to any external knowledge base. Its use is largely limited to specialist applications such as trademark registration, identification of drawings in a design archive, or colour matching of fashion accessories.

Level 2 comprises retrieval by derived (sometimes known as logical) features, involving some degree of logical inference about the identity of the objects depicted in the image. It can usefully be divided further into:
1. retrieval of objects of a given type (e.g. “find pictures of a double-decker bus”);
2. retrieval of individual objects or persons (“find a picture of the Eiffel tower”).

To answer queries at this level, reference to some outside store of knowledge is normally required – particularly for the more specific queries at level 2(b). In the first example above, some prior understanding is necessary to identify an object as a bus rather than a lorry; in the second example, one needs the knowledge that a given individual structure has been given the name “the Eiffel tower”. Search criteria at this level, particularly at level, are usually still reasonably objective.

Level 3 comprises retrieval by abstract attributes, involving a significant amount of high-level reasoning about the meaning and purpose of the objects or scenes depicted. Again, this level of retrieval can usefully be subdivided into:
1. retrieval of named events or types of activity (e.g. “find pictures of Scottish folk dancing”);
2. retrieval of pictures with emotional or religious significance (“find a picture depicting suffering”).

Success in answering queries at this level can require some sophistication on the part of the searcher. Complex reasoning, and often subjective judgement, can be required to make the link between image content and the abstract concepts it is required to illustrate. Queries at this level, though perhaps less common than level 2, are often encountered in both newspaper and art libraries.

Figure 4: Architecture of CBIR

6 IMPLEMENTATION DETAILS

6.1 Color Histogram Calculation

We define color histograms as a set of bins where each bin denotes the probability of pixels in the image being
of a particular color. A color histogram $H$ for a given image is defined as a vector:

$$H = \{H[0], H[1], H[2], \ldots \ldots H[i], \ldots H[n]\}$$

Where ‘$i$’ represents color in the color histogram[i] is the Number of pixels in color ‘$i$’ in that image and ‘$n$’ is the number of bins in the color histogram. Typically, each pixel in an image will be assigned to a bin of a color histogram of that image, so for the color histogram of an image, the value of each bin is the number of pixels that has the same corresponding color [2]. In order to compare images of different sizes, color histograms should be normalized. The normalized color histogram ‘$H’$ is defined as: $H’ = \{H’[0], H’[1], H’[2], \ldots \ldots H’[i], \ldots H’[n]\}$. Where $H’[i]$ is $H’[i]/p$, $p$ is the total number of pixels in an image. An ideal color space quantization presumes that distinct colors should not be located in the same sub-cube and similar colors should be assigned to the same sub-cube. Using few colors will decrease the possibility that similar colors are assigned to different bins, but it increases the possibility that distinct colors are assigned to the same bins, and that the information content of the images will decrease by a greater degree as well. On the other hand, color histograms with a large number of bins will contain more information about the content of images, thus decreasing the possibility of distinct colors will be assigned to the same bins. However, they increase the possibility that similar colors will be assigned to different bins, the storage space of metadata, and the time for calculating the distance between color histograms. Therefore, there is a trade-off in determining how many bins should be used in color histograms.

6.2 Converting RGB to HSV

The value is given by

$$V = \frac{1}{2} \left(1 - \frac{3}{2} \frac{\min(R, G, B)}{R + G + B}\right)$$

$$S = 1 - \frac{3}{2} \frac{\min(R, G, B)}{R + G + B}$$

$$H = \cos^{-1}\left(\frac{\frac{1}{2} \left(\frac{(R - G) + (R - B)}{(R - G)^2 + (R - B)(G - B)}\right)^\frac{1}{2}}{V}\right)$$

6.3 Edge Histogram Calculation

Edges in images constitute an important feature to represent their content. Also, human eyes are sensitive to edge features for image perception. One way of representing such an important edge feature is to use a histogram. An edge histogram in the image space represents the frequency and the directionality of the brightness changes in the image. It is a unique feature for images. To represent this unique feature, in MPEG-7, there is a descriptor for edge distribution in the image. This Edge Histogram Descriptor (EHD) proposed for MPEG-7 expresses only the local edge distribution in the image. That is, since it is important to keep the size of the descriptor as compact as possible for efficient storage of the metadata, the MPEG-7 edge histogram is designed to contain only 80 bins describing the local edge distribution. These 80 histogram bins are the only standardized semantics for the MPEG-7 EHD [1]. However, using the local histogram bins only may not be sufficient to represent global features of the edge
distribution. Thus, to improve the retrieval performance, we need global edge distribution as well.

The EHD basically represents the distribution of 5 types of edges in each local area called a sub-image. As shown in Fig. 5, the sub-image is defined by dividing the image space into 4x4 no overlapping blocks. Thus, the image partition always yields 16 equal-sized sub-images regardless of the size of the original image. To characterize the sub-image, we then generate a histogram of edge distribution for each sub-image. Edges in the sub-images are categorized into 5 types: vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional (Fig. 6). Thus, the histograms for each sub-image represent the relative frequency of occurrence of the 5 types of edges in the corresponding sub-image.

As a result, as shown in Fig. 6, each local histogram contains 5 bins. Each bin corresponds to one of 5 edge types. Since there are 16 sub-images in the image, a total of 5x16=80 histogram bins is required. Note that each of the 80-histogram bins has its own semantics in terms of location and edge type. The semantics of the histogram bins form the normative part of the MPEG-7 standard descriptor. Specifically, starting from the sub-image at (0,0) and ending at (3,3), 16 sub-images are visited in the raster scan order and corresponding local histogram bins are arranged accordingly [16]. Within each sub image, the edge types are arranged in the following order: vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional. Table 1 summarizes the complete semantics for the EHD with 80 histogram bins.

Of course, each histogram bin value should be normalized and quantized. For normalization, the number of edge occurrences for each bin is divided by the total number of image-blocks in the sub-image.

Image block is a basic unit for extracting the edge information. That is, for each image-block, we determine whether there is at least an edge and which edge is predominant. When an edge exists, the predominant edge type among the 5 edge categories is also determined. Then, the histogram value of the corresponding edge bin increases by one. Otherwise, for the monotone region in the image, the image-block contains no edge. In this case,
that particular image-block does not contribute to any of the 5 edge bins. Consequently, each image-block is classified into one of the 5 types of edge blocks or a non edge block. Although the non edge blocks do not contribute to any histogram bins, each histogram bin value is normalized by the total number of image-blocks including the non edge blocks. This implies that the summation of all histogram bin values for each sub-image is less than or equal to 1. This in turn, implies that the information regarding non edge distribution in the sub-image (smoothness) is also indirectly considered in the EHD. Now, the normalized bin values are quantized for binary representation. Since most of the values are concentrated within a small range (say, from 0 to 0.3), they are nonlinearly quantized to minimize the overall number of bits. Since the EHD describes the distribution of non-directional edges and non edge cases as well as four directional edges, the edge extraction scheme should be based on the image-block as a basic unit for edge extraction rather than on the pixel. That is, to extract directional edge features, we need to define small square image-blocks in each sub-image as shown in Fig. 5. Specifically, we divide the image space into non overlapping square image-blocks and then extract the edge information from them.

Note that, regardless of the image size, we divide the image space into a fixed number of image-blocks. The purpose of fixing the number of image-blocks is to cope with the different sizes (resolutions) of the images. That is, by fixing the number-of-image blocks, the size of the image block becomes variable and is proportional to the size of the whole image. The size of the image-block is assumed to be a multiple of 2. Thus, it is sometimes necessary to ignore the outmost pixels in the image to satisfy that condition. Simple method to extract an edge feature in the image block is to apply digital filters in the spatial domain. To this end, we first divide the image-block into four sub-blocks. Then, by assigning labels for four sub-blocks from 0 to 3, we can represent the average gray levels for four sub-blocks at (i,j)th image-block as $a_0(i,j), a_1(i,j), a_2(i,j)$, and $a_3(i,j)$, respectively.

$$m_e(i, j) = \sum_{k=0}^{3} a_k(i,j) \times f_e(k) \quad [1]$$

$$m_v(i, j) = \sum_{k=0}^{3} a_k(i,j) \times f_v(k) \quad [2]$$

$$m_{d-45}(i, j) = \sum_{k=0}^{3} a_k(i,j) \times f_{d-45}(k) \quad [3]$$

$$m_{d+45}(i, j) = \sum_{k=0}^{3} a_k(i,j) \times f_{d+45}(k) \quad [4]$$

$$m_{d+135}(i, j) = \sum_{k=0}^{3} a_k(i,j) \times f_{d+135}(k) \quad [5]$$

Also, we can represent the filter coefficients for vertical, horizontal, 45-degree diagonal, 135-degree diagonal, and non-directional edges as $f_v(k), f_h(k), f_{d+45}(k), f_{d-135}(k)$, and $f_{nd}(k)$, respectively, where $k=0,1,2,3$ represents the location of the sub-blocks. Now, the respective edge
magnitudes \( m_{v(i,j)}, m_{h(i,j)}, m_{d-45(i,j)}, m_{d-135(i,j)} \), and \( m_{nd(i,j)} \) for the \((i,j)\)th image-block can be obtained as follows:

\[
\begin{align*}
&\text{(a) vertical} & & \text{(b) horizontal} \\
&m_{v(i,j)} & & m_{h(i,j)} \\
&m_{d-45(i,j)} & & m_{d-135(i,j)} \\
&m_{nd(i,j)} & & \\
\end{align*}
\]

Figure 8: Filter Coefficients For Edge Detection

The maximum value among 5 edge strengths obtained from (1) to (5) in the image-block is considered to have the corresponding edge in it. Otherwise, the image-block contains no edge.

\[
\max\{m_{v(i,j)}, m_{h(i,j)}, m_{d-45(i,j)}, m_{d-135(i,j)}, m_{nd(i,j)}\} = \max\{m_{v(i,j)}, m_{h(i,j)}, m_{d-45(i,j)}, m_{d-135(i,j)}, m_{nd(i,j)}\} 
\]

6.4 Distance Measures

We tried 3 kinds of histogram distance measures for a histogram \( H(i), i=1,2,...,N \)

6.4.1 L-2 Distance

Defined as:

\[
d_{L-2}(q,i) = \left( \sum_{m=1}^{N} (h_q(m) - h_i(m))^2 \right)^{1/2}
\]

this Metric is uniform in terms of the Euclidean distance between vectors in feature space, but the vectors are not normalized to unit length.

6.4.2 Cosine Distance

If we normalize all vectors to unit length, and look at the angle between them, we have cosine distance, defined as:

\[
d_{cos}(q,i) = \frac{2}{\pi} \cos^{-1} \left( \sum_{m=1}^{N} \frac{h_q(m)h_i(m)}{\min(h_q[m],h_i[m])} \right)
\]

6.4.3 Histogram Intersection

Defined as:

\[
d_{int}'(q,i) = 1 - \frac{\sum_{m=0}^{M-1} \min(h_q[m],h_i[m])}{\min(\|h_q\|,\|h_i\|)}
\]

The denominator term is needed for non-normalized histogram features

6.5 Combining Features and Making Decisions

This section formulates the problem of combining different features as a problem of finding a set of weights of different feature distances, and then presents a Mini-Max algorithm in finding the best-matching image, Mini-Max Combination. Let’s first look at a more general case, where we have: query image \( q \), images in the database \( i \), \( K \) features and thus \( K \) kinds of distances (they are constants) \( d_k(q,i), k=1,...,K, i=1,...,N \)

Assume we are going to combine them as a weighted sum of all the distances, i.e. the distance for an image in the database is written as:

\[
D(q,i) = \sum_{k=1}^{K} w_k d_k(q,i)
\]

Now we want to search for a vector \( w \) that satisfies Eq.(2) and the resulting distance measure is “most close” to our subjective criteria. There are two candidate approaches:

i) assign a set of weights based on the perceptually judgment of the designer on some image set (training).

But the problem here is that this set of weights may perform poorly on new dataset.

ii) Or, having no assumption about the subjective judgment of a user, we choose the image that minimizes the maximum distance over all valid set of weights as the best match (denoted as Mini-Max hereafter). For every image \( i \), searching for the maximum distance over the weight space turns out to be a linear program, a thus have fast solution:
Maximize: \((1)\), Subject to \((2)\). Where all \(d_s\) are the constants and \(w_k, k = 1, ..., K\) are unknown. The image with the minimum “max-distance” is declared as the best match to the query image. For our 2 features case the max distance
\[ D(q, i) = w_d d_d(q, i) + (1 - w) d_t(q, i), 0 \leq w \leq 1 \]
of every image \(i\), is a linear function of \(w\) over \([0, 1]\). Thus the maximum either lies at \(w=0\) or \(w=1\), and comparing \(d_d(q, i)\) and \(d_t(q, i)\) is sufficient. Then we rank the maximum of \(d_d(q, i)\) and \(d_t(q, i)\) for all \(i\), and take \(n\) images with the least distance as our return result.

7. Testing

7.1 Recall and Precision Evaluation

Testing the effectiveness of the content based image Retrieval about testing how well the CBIR can retrieve similar images to the query image and how well the system prevents the return results that are not relevant to the source at all in the user point of view. A sample query image must be selected from one of the image category in the database. When the system is run and the result images are returned, the user needs to count how many images are returned and how many of the returned images are similar to the query image. Determining whether or not two images are similar is purely up to the user’s perception. Human perceptions can easily recognise the similarity between two images although in some cases, different users can give different opinions. After images are retrieved, the system’s effectiveness needs to be determined. To achieve this, two evaluation measures are used. The first measure is called Recall. It is a measure of the ability of a system to present all relevant items. The equation for calculating recall is given below:

\[ \text{Recall} = \frac{\text{Number of relevant items Retrieved}}{\text{Number of relevant items in Collection}} \]

The second measure is called Precision. It is a measure of the ability of a system to present only relevant items. The equation for calculating precision is given below:

\[ \text{Precision} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of items retrieved}} \]

The number of relevant items retrieved is the number of the returned images that are similar to the query image in this case. The number of relevant items in collection is the number of images that are in the same particular category with the query image. The total number of items retrieved is the number of images that are returned by the system.

7.2 Test Cases

The CBIR implementation consists of various subsystems for color and texture that are built out of modules, which are composed of procedures and functions. The testing process hence consists of different stages, where testing is carried out incrementally in conjunction with system implementation. We have conducted several color and texture experiments on the test database of 250 images. The sample results of these experiments are given as follows:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>29</td>
<td>45</td>
<td>63.04</td>
<td>64.44</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>30</td>
<td>44</td>
<td>78.94</td>
<td>68.00</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>12</td>
<td>20</td>
<td>66.66</td>
<td>60.00</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>25</td>
<td>36</td>
<td>83.33</td>
<td>72.22</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>23</td>
<td>35</td>
<td>59.37</td>
<td>65.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>19</td>
<td>45</td>
<td>41.30</td>
<td>42.22</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>28</td>
<td>44</td>
<td>73.33</td>
<td>63.64</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>11</td>
<td>20</td>
<td>61.11</td>
<td>55.00</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>22</td>
<td>36</td>
<td>73.33</td>
<td>61.11</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>19</td>
<td>35</td>
<td>59.37</td>
<td>54.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>35</td>
<td>45</td>
<td>76.08</td>
<td>77.77</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>33</td>
<td>44</td>
<td>86.84</td>
<td>75.00</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>15</td>
<td>20</td>
<td>83.33</td>
<td>75.00</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>26</td>
<td>36</td>
<td>86.67</td>
<td>72.22</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>27</td>
<td>35</td>
<td>84.38</td>
<td>77.14</td>
</tr>
</tbody>
</table>
Consider the Column names of the tables as A= Test; B= Total Number of Relevant Images; C = Number of relevant images Retrieved; D = Total Number of Retrieved Images; E = Recall; F = Precision.

8. CONCLUSION

The dramatic rise in the sizes of images databases has stirred the development of effective and efficient retrieval systems. The development of these systems started with retrieving images using textual annotations but later introduced image retrieval based on content. This came to be known as CBIR or Content Based Image Retrieval. Systems, using CBIR we can retrieve images based on visual features such as colour, texture and shape, as opposed to depending on image descriptions or textual indexing. In this paper we proposed an image retrieval system that evaluates the similarity of each image in its data store to a query image in terms of colour and textural characteristics, and returns the images within a desired range of similarity. From among the existing approaches to color and texture analysis within the domain of image processing, we have adopted the MPEG-7 edge histogram and color histogram to extract texture and color features from both the query images and the images of the data store. For distance measuring between histogram vectors of two images, we have tried three distance measures; they are l2 distance measure, cosine distance measure and histogram intersection distance measure. The experiment results showing that, the average relevant image retrieval rate is increased by 10% by combined features of color and texture than considering features individually.

REFERENCES


[3] Dr. Fuhui long, Dr. Hongjiang and Prof. David Dagan Feng, “Fundamentals of content based image retrieval”


Author’s Biography

Srinivasa Kumar Devireddy received the B.E. degree in Computer Science & Engineering from Karnataka University, Dharwad in 1992 and M.S. degree in Software Systems from Birla Institute of Technology and Science, Pilani in 1995. He is currently working as Principal, St. Mary’s Women’s Engineering College, Guntur, A.P., and India. He is a member of IEEE. His research interests are in the areas of Biometrics, Image Processing and Content Based Image Retrieval.
Video Decoding In Spatial Scalability Using Inter Layer Prediction

Krishna Reddy, K.R. 1 Nagabhushana 2 Aravinda T.V. 3 G.Mahadevan 4

1,2,3 Ph.D. Student, PRIST University, Thanjavur, Tamil Nadu. Email: skishu2001@yahoo.co.in, nagabhushana069@yahoo.com, arvind_cta@yahoo.co.in
2Prof.&Head, Dept. of CS&E, AMCE, Bangalore, Karnataka. Email: g_mahadevan@yahoo.com

1733

ABSTRACT
The scalability extension of H.264/AVC uses an oversampled pyramid representation for spatial scalability, where for each spatial resolution a separate motion compensation or MCTF loop is deployed. When the reconstructed signal at a lower resolution is used to predict the next higher resolution, the motion compensation or MCTF loops including the deblocking filter operations of both resolutions have to be executed. This imposes a large complexity burden on the decoding of the higher resolution signals, especially when multiple spatial layers are utilized. In this paper, we investigate the approach to only allow prediction between spatial layers for parts of the lower resolution pictures that are intra-coded in order to avoid decoding that requires multiple motion compensation or MCTF loops. Experimental results evaluate the effectiveness of the proposed approach.

Keywords: Scalability, H.264/AVC, Inter-Layer Prediction, Single-loop decoding.

1. INTRODUCTION
The scalable extension of H.264/AVC as proposed in [1][2][3][13] has been chosen to be the starting point of MPEG’s Scalable Video Coding (SVC) project in October 2004. In January 2005, the ISO/IEC Moving Pictures Experts Group (MPEG) and the Video Coding Experts Group (VCEG) of the ITU-T agreed to jointly finalize the SVC project as an Amendment of their H.264/AVC standard [3][4][10], and the scalable extension of H.264/AVC was selected as the first Working Draft [5]. A reference encoder is described in the Joint Scalable Video Model 0 (JSVM 0) [6].

H.264/AVC is a hybrid video codec specifying for macroblocks either motion-compensated prediction or intra prediction. Both predictions are followed by residual coding. The basic design of the scalable extension of H.264/AVC can be classified as layered video codec. In each layer, the basic concepts of motion-compensated prediction and intra prediction are employed as in standard H.264/AVC. However, additional inter-layer prediction mechanisms have been integrated in order to exploit the redundancy between several layers. SNR scalability is basically achieved by residual quantization with little changes to H.264/AVC. For spatial scalability, a combination of motion-compensated prediction and oversampled pyramid decomposition is proposed, which requires some additional mechanisms to convey bit-rate from the lower resolution to the higher resolution layers. Because of the similarities in motion-compensated prediction, the approach to temporal scalability of H.264/AVC is maintained [9][11].

Among the various types of scalability, spatial scalability requires the largest degree of change to H.264/AVC. Various mechanisms to re-use information from a lower spatial resolution in a higher spatial resolution layer are
specified. One of these mechanisms is the upsampling of the decoded signal at the lower resolution and making this signal available for prediction. However, this feature has the drawback that both motion compensation loops of the lower and higher resolution must be executed for parts of the picture that are not intra-coded in the lower resolution signal. In this paper, we investigate an approach that enables a single-loop decoding by constraining the inter-layer prediction.

The next section outlines the basic concepts of the scalability extension of H.264/AVC while a more detailed description can be found in [1][2][5][16]. Section III describes the constrained inter-layer prediction for enabling single-loop decoding, and Section IV provides experimental results comparing single-loop and multiple-loop decoding.

2. SCALABLE EXTENSION OF H.264/AVC

The scalable H.264/AVC extension specifies a layered video codec. In general, the coder structure depends on the scalability space that is required by the application. For illustration, Fig. 1 shows a typical coder structure with 2 spatial layers. In each layer, which can either be a spatial layer or a coarse-grain SNR layer, an independent hierarchical motion-compensated prediction structure with layer-specific motion parameters is employed. The redundancy between consecutive layers is exploited by different inter-layer prediction concepts that include prediction mechanisms for motion parameters as well as texture data.

A base representation of the input pictures of each layer is obtained by transform coding similar to that of H.264/AVC, the corresponding NAL units (NAL – Network Abstraction Layer) contain motion information and texture data; the NAL units of the base representation of the lowest layer are compatible with standard H.264/AVC. The reconstruction quality of the base representations can be improved by an additional coding of so-called progressive refinement slices; the corresponding NAL can be arbitrarily truncated in order to support fine granular quality scalability (FGS) or flexible bit-rate adaptation.
residuals (dashed arrows in Fig. 2) are introduced in addition to the motion compensated prediction. For more details on how the hybrid coding approach of H.264/AVC is extended towards MCTF please refer to [1][2][5][6][11][14].

B. SNR Scalability
For the SNR base layer (base representation), H.264/AVC conforming transform coding is used [15]. For each macroblock, the coded block pattern (CBP), and conditioned on CBP the corresponding residual blocks are transmitted together with the macroblocks modes, intra prediction modes, and motion data using the I, P, or B slice syntax of H.264/AVC.

On top of the SNR base layer, SNR enhancement layers are coded. For that, the quantization error between the SNR base layer and the original residual and intra macroblocks is re-quantized exactly using the same methods as for the base layer but with a finer quantization step size, i.e., a lower value of the quantization parameter. In a simple version, the transform coefficient levels of the SNR enhancement layers are transmitted using the residual syntax of H.264/AVC.

Note that, it is basically also possible to specify motion field refinements for SNR enhancement layers. Therefore, the same inter-layer prediction techniques as described in the next subsection but without the upsampling operations are applied.

C. Inter-Layer Prediction and Spatial Scalability
As a first interpretation, the pictures (base representations) for different layers are coded independently with layer-specific motion information. We consider spatial scalability with a factor of 2 in horizontal and vertical resolution [12], although the concepts can be generalized.

From several experiments we have found that it would be efficient to allow the encoder to freely choose which dependencies between spatial resolution layers need to be exploited through switchable prediction mechanisms. The following techniques turned out to provide gains and were included into the scalable video codec:

- Prediction of motion vectors using the upsampled lower resolution motion vectors.
- Prediction of the residual signal using the upsampled residual signal of the lower resolution layer.
- Prediction of a macroblock using the reconstructed and upsampled lower resolution signal.

The last of the 3 methods is the one modified in this work while the other two remain unchanged. The inter-layer prediction techniques are briefly described in the following.

1) Motion Vector Prediction: For prediction of motion vectors using the upsampled lower resolution motion vectors we have introduced two additional macroblock modes that utilize motion information of the lower resolution layer. The macroblock partitioning is obtained by upsampling the partitioning of the corresponding 8x8 block of the lower resolution layer. For the obtained
macroblock partitions, the same reference picture indices as for the corresponding sub macroblock partition of the base layer block are used; and the associated motion vectors are scaled by a factor of 2. While for the first of these macroblock modes no additional motion information is coded, for the second one, a quarter-sample motion vector refinement is transmitted for each motion vector. Additionally, our approach includes the possibility to use a scaled motion vector of the lower resolution as motion vector predictor for the conventional motion-compensated macroblock modes.

2) Residual Prediction: In order to also incorporate the possibility of exploiting the residual information coded in the lower resolution layer, an additional flag is transmitted for each macroblock, which signals the application of residual signal prediction from the lower resolution layer. If the flag is true, the base layer residual signals is block-wise up-sampled using a bi-linear filter with constant border extension and used as prediction for the residual signal of the current layer, so that only the corresponding difference signal is coded.

3) Intra Prediction: We have further introduced an additional intra macroblock mode. In that mode, the intra prediction signal is generated by upsampling the reconstruction signal of the lower resolution layer using the 6-tap filter which is specified in H.264/AVC for the purpose of half-sample interpolation. The prediction residual is transmitted using H.264/AVC residual coding.

3. CONSTRAINED INTER-LAYER PREDICTION

For the inter-layer prediction using the reconstructed lower resolution signal as described in Sec. II.C-3 it is generally required that the lower resolution layer is completely decoded including the computationally complex operations of motion compensated prediction (or inverse MCTF) and de-blocking. The blue-marked pictures represent the upsampled versions of the decoded layer $k$ pictures. For decoding layer $k+1$, the orange-marked macroblocks are predicted from the decoded and upsampled pictures of layer $k$. It is worth noting that at the decoder only those parts of layer $k$ need to be upsampled that are actually used for prediction. However, the main problem remains that in general the motion compensated prediction (or inverse MCTF) as well as the de-blocking for layer $k$ must be executed to decode layer $k+1$. This creates a large complexity overhead for the decoding process of layer $k+1$ and even more for all higher layers.

We have found that the above problem can be circumvented by restricting the prediction from upsampled decoded pictures to those parts of the lower layer picture that are coded with intra macroblocks. For that, the intra prediction signal is directly obtained by

![Figure 4: Padding of Intra Macroblocks Before Upsampling](image)

The interpolation of an 8x8 block of the lower layer is generally performed using the half-pel interpolation filter of H.264/AVC. Before interpolation, the block edges
inside intra macroblocks as well as the macroblock edges between intra macroblocks of the base layer are deblocked as specified in H.264/AVC, and afterwards these modified intra macroblocks are extended by a 4-pixel border in each direction using the following padding process (see Fig. 4).

When a neighboring macroblock is coded in an intra mode no border extension is performed but the corresponding intra samples are used for interpolation (Fig. 4b,c,d). Otherwise, the 4-pixel border is generally obtained by horizontal constant border extension of the current macroblock (as well as of vertical neighbors coded in intra mode) and a subsequent vertical border extension. However, if a horizontal or vertical neighboring macroblock is coded in an inter mode but one of the two diagonal neighboring macroblocks is coded in intra mode, then the 4x4 block of the border extension that is located next to the diagonal intra-coded neighbor (e.g. magenta colored block in Fig. 4c) is generated similar to the intra prediction signal for the diagonal down right intra prediction mode of H.264/AVC. When the corresponding 4x4 block does not represent the upper block of the right border, the coordinates that are used in the intra prediction process are modified accordingly. If the corner sample that is needed for generating the intra prediction signal does not belong to an intra coded macroblock as in Fig. 4d, it is replaced by the average of the two neighboring corner samples.

4. EXPERIMENTAL RESULTS

For evaluating the impact of the proposed constrained inter-layer prediction on coding efficiency, we compared it with the general scheme [1][16] that requires a multiple-loop decoding. Note that multiple-loop decoding stands for the unrestricted version of prediction of a macroblock using the reconstructed and upsampled lower resolution signal. We have chosen a scalability scenario that consists of 5 spatial or coarse-grain SNR layers. The lowest layer is coded conforming to the H.264/AVC standard; for all enhancement layers MCTF has been applied. With the general multiple-loop scheme, 5 motion compensation or MCTF loops are required for decoding the highest layer; whereas with the proposed modification, the decoding of the highest layer can be realized with a single MCTF loop. For both codec versions the same encoder control following [8] was used.

Our simulation results have shown that for most sequences, the impact on coding efficiency is small by imposing our proposed restriction. An example for such a case which stands for a larger number of other cases is shown in Fig. 5a for the Foreman sequence. In addition to the many cases for which we have found small impairments due to our proposed restriction leading to single-loop decoding, we have found two sequences, namely Football and Crew, for which we have observed a more significant influence on the coding efficiency. As an example, the result for Football is depicted in Fig. 5b.

In Fig. 6, the usage of macroblock modes for the first CIF 15Hz layer has been analyzed for the sequences Foreman and Football. “Intra” represents the intra modes of standard H.264/AVC, while “Intra_BL” stands for the
intra mode that uses the upsampled base layer signal as prediction. “BLMode” stands for the macroblock modes that employ the motion partitioning, the reference indices, and motion vectors of the base layer (cp. Sec. II.C-1), while the motion-compensated macroblock modes of standard H.264/AVC are labeled by the term “NormalMC”. The suffix “+RP” indicates the additional usage of inter-layer residual prediction (cp. Sec. II.C-2) for the motion-compensated macroblock modes.

The analysis shows that the largest impairments of the coding efficiency are observed for sequences, for which a large degree of macroblocks (about 27% for the Football sequence) are coded using the “Intra_BL” mode with the general unrestricted inter-layer prediction scheme. By constraining the usage of the “Intra_BL” mode, a major part of these macroblocks is inter-coded, mainly with additional residual significant impairment of the coding efficiency is observed, are mainly characterized by fast and complex motion that cannot be well presented by motion-compensated prediction.

5. Conclusion
We have presented a simple modification to the inter-layer prediction in pyramid-based spatial scalability. The approach is integrated into the scalability extension of H.264/AVC which was chosen as the first Working Draft of the new JVT standardization activity on Scalable Video Coding. By restricting the prediction to upsampled intra-coded image parts, a single-loop decoder for spatial and coarse-grain SNR scalability can be realized requiring only the motion compensated prediction or inverse MCTF as well as deblocking of inter-coded macroblocks for the scalable layer that is being decoded. The rate-distortion penalty for this restriction is found for most sequences to be small while only a few sequences are found with PSNR losses up to 0.7 dB.

References
Video Decoding In Spatial Scalability Using Inter Layer Prediction


Author’s Biography

Dr. G. Mahadevan M.E., Ph.D. (CSE) is having around 20 years of experience in teaching field. He has 4 years of experience, working as Principal of A.R.J College of Engg & Tech, Mannargudi, Thanjavur. Presently, he is on the Board of Examiners and is in LIC committee for VTU Belgaum. Currently, he is guiding 4 Ph.D candidates for Anna University Coimbatore and 3 from PRIST University Thanjavur. He is doing consultancies for 15 companies. He has published more than 26 national and international papers. He conducted more than 6 international and 8 national conferences. He has received government funds for research programs. He has completed B.E from Mookambika college of Engineering Trichy in 1989.
completed M.Tech from REC, Trichy in 1994. He completed his Ph.D in CSE in the year 2002

Krishna Reddy K.R., BE,M.Tech, (Ph.D) (CSE), is having around 14 years of experience in teaching field. Presently working as a Assistant Professor in Computer science & engineering department, SJM Institute of Technology, chitradurga. Currently he is pursuing Ph.D. (CSE) in PRIST University, Thanjavur. He has published 2 papers in national level journals. He has completed BE from Mysore university and M.tech from KREC, Mangalore university.

Nagabhushana, BE (CS&E), M.Tech, (Ph.D) (CSE), is having around 7 years of experience in teaching field. Presently working as a Sr. lecturer in Information science & engineering department, S.J.M Institute of Technology, Chitradurga. Currently he is pursuing Ph.D. (CSE) in PRIST University, Thanjavur. He has published 2 national level papers. He has completed BE (CSE) from Kuvempu University and M.Tech from AAI Deemed University.

Aravinda.T.V, BE (CS&E), M.Tech, (Ph.D) (CSE), is having around 7 years of experience in teaching field. Presently working as lecturer in Computer science & engineering department, GM Institute of Technology, Davangere. Currently he is pursuing Ph.D. (CSE) in PRIST University, Thanjavur. He has published 2 national level papers. He has completed BE (CSE) from Kuvempu University, and M.Tech from AAI Deemed University.
Detection, Counting and Classification of Moving Objects by Using Real Time Traffic Flux through Differential and Rule Based Analysis

H.S. Mohana1          M. Aswathakumar2

1Malnad College of Engineering, Hassan, Karnataka
2MS Ramaiah Institute of Technology, Bangalore, Karnataka. Email : hsm@mcehassan.ac.in, maswatha@yahoo.com

ABSTRACT
This paper is focussed on detection, counting and classification of moving objects using differential and graphical techniques. The basic idea used is variation in the traffic flux density due to presence of objects in the scene. Accurate Traffic flux estimation will play vital role in object detection, counting and classification. The designed technique is evaluated with 15 different video sequences and weighed thoroughly with simple confidence measures. In the present work we have achieved real time analysis with normal video rate. And for object classification computation we are taking specific frame gap which saves computational time. The result produced with this analysis is extremely good and beneficial in real time traffic control, detecting and classifying objects in urban areas. The vehicle counting achieved with an accuracy of 94% under varied road conditions. In the normal condition the average accuracy achieved is near to 97%.

Keywords: Traffic Flux, Object Detection, Objects Classification, Differential and Graphical Techniques, Dynamic Selection, Ego-Motion.

1. INTRODUCTION
Most of the motion picture analysis presently available, takes considerable computational time, although we have high-speed computational technology. Real time view analysis will be very challenging as it involves the time component. Here, in this work an attempt is made to introduce a robust, simple and statistical solution to this difficult problem. Two stages of solution have been designed, tested and implemented in the present work. Firstly, to reduce the numbers of frames used for analysis, dynamic selection of images were made. Here, the frame-to-frame difference is obtained and a threshold has been fixed to register a subset of images to be used for analysis out of the 15/30 frames available for every second. This results considerable saving of computation and time. The selected subset is compared with the reference template, which is selected background image. This is done under different illumination conditions. In the second phase of the work, reference frame is constantly subtracted from the dynamically selected subset. This leads to the separation of object pixel, which is corresponding to moving vehicles and the background pixel, which are not altered. Counting object pixel in a frame leads to the traffic flux estimation. To make the design illumination invariant, a section of the background is taken as reference, which will not be affected by the traffic flow. Comparing the illumination of that block of reference with present picture will decide which background reference must be considered, for the purpose of analysis. Since it is small matrix pixel, time constraints of computation have been tackled. This novel and simple statistical algorithm is tested over real image sequence. Discrimination of object pixel and background pixel has shown good repeatability.
over many real sequences of images. Threshold is fixed and used to discriminate the low, medium and high traffic flux. There is a plot for traffic flux density; it’s basically % flux density versus number of frames. The vehicle detection and classification is carried out by using this plot. Suppose if there is any object in the scene, then there is a certain range of flux density will produce according to object volume (size). Obviously the object pixel count is directly reflecting the presence of vehicles as well as type. By analyzing flux density in various cases, we can classify the objects (pedestrian, two wheeler, four wheeler and heavy vehicles).

Traffic flux is a generic term and it is nothing but the change in object pixel against the frame number. This embeds enormous information about the moving objects in the scene. By analyzing this percentage flux density variation, we can classify the objects. Hence, accurate Traffic flux estimation will play a very vital role in object detection and classification [14, 15]. The presence of object in the scene contributes a definite increase in flux. Similarly, exit of object from the scene decreases the flux. The plot of the same results in a stochastic variation graph. Presence of different vehicles/object in the scene creates different patterns in terms of change in percentage flux and slope. This forms the clear cut basis for the analysis of moving object/vehicles, counting and classification of the same. Object classification is done by comparing percentage flux density, if there is only one vehicle at a time in the scene. Suppose if there are more objects at a time in the scene, in this case we have to analyze percentage flux density to classify the objects and we have methodology for classification. Suppose if there are multiple vehicles simultaneously enter into the scene still contribution by each of them are different along with different exit time. However, the accuracy of detection, counting and classification is cent percent when a single object or series of objects present in the video stream. The designed technique is evaluated with 15 different video sequences. The results obtained are encouraging and also establishes that even multiple entry and exit does not result in largely reduced accuracy. Very good result is achieved by combining logical statements with multiple threshold values. The module being much generalized, therefore, it fits into the applications of various areas, such as detecting speed violators on highways, vehicle count, weather forecasting, cloud tracking, satellite image processing for surveillance, biomedical analysis and so on.

2. Related Work
The review of the literature pertaining to the present topic is presented to the readers. In [1] authors worked on comparison of different approaches of optical flow estimation. Comparison is done on the basis of accuracy and computational complexities. They have concluded differential technique is best suited for the competition of optical flow and hence the dynamic scene analysis. Entropy based features are used in [2], to check for the existence of vehicles and then tracking is achieved. Though this takes less computational time it suffers serious occlusion problem. Fusion of images and vector maps technique is used in [3] to discriminate vehicles from objects in the scene. This is suitable for military applications as overall system is complicated and expensive. A comparison of edge element association EEA and marginalized contour approaches for 3D model based vehicle tracking in traffic scenes is implement in [4]. Tracking failures of two approaches, however, usually do not happen at the same time frames which can lead to insights into relative strengths and weakness of the two approaches. Since both the models are to be implemented on every frame computational time frame increases. Recursive optical flow estimation- Adaptive filtering
Detection, Counting and Classification of Moving Objects by Using Real Time Traffic Flux through Differential and Rule Based Analysis

approach is used in [5]. This is modification over Horn and Schunck[6] algorithms as it uses only parts of images. Hence sequence of images is used with adaptive filtering technique. The result achieved here is good at cost of linearly growing computational complexities because convergence to be achieved. In [7] authors present a new method for tracking rigid objects using a modified version of the Active Appearance Model. It works well with partial and self occlusion of objects. The layered representation is more flexible than standard image transforms and can capture many important properties of natural image sequences [8]. The study reveals that increased computational time and complexities are the hurdles in achieving real time analysis at video rate. This fact motivated us to develop the simple technique presented in this paper. The algorithm developed in this work computes the simple differences required and avoids time consuming iterative part in the optical flow analysis. Wherever possible certain parameters are evaluated and supported with sufficient sample sequences. Result establishes that they work well with the inherent problem of natural video traffic sequences. This paper emphasizes on the time and computational complexities of the developed simple algorithms, as there is a need of estimating the traffic flux in real time. In the present work we are concentrating on the accuracy of vehicle count and classification. In the end, it is important to ascertain the conditions under which these results were obtained. Traffic flux estimation critically depends on the changes in the intensities of the I^th image with respect to the reference image at all spatially uniformly spread pixels. One of our assumptions is that the intensities of the moving vehicles are preserved during the travel in the view path.

2.1 General Motion
In general, an observed motion does not have the simple structure of the spatially constant motion as assumed. Although motions are not constant in space still can make sense. Despite of different processing algorithms, three stages processing is essential to perform computing the motion in spatio-temporal domain.

1. Pre-filtering or smoothing with low-pass or band pass filters in order to extract signal structures of interest and to enhance the signal to noise ratio.
2. The extraction of basic measurements, such as spatio-temporal derivatives or local correlation surfaces
3. The integration of these measurements to produce 2D flow field, which often involves assumptions about the smoothness of the underlying flow field.

The algorithm description and analysis assumes an affined camera where perspective effects are limited to changes in overall scale. No camera calibration parameters are required since the assumptions are made as mentioned before. Camera used is of resolution 1024 × 1024 with a video rate of 30 frames per second. To reduce the time of computations, the same has the resolution of the image is scaled down to 200 × 200 without losing much of the information. The above size reduction saves computation.

2.2 Differential Methods
Differential techniques compute motion related information from spatio-temporal derivatives of image intensity. The differential technique developed by Horn and Schunk[6] has been the most widely used algorithm for the optical flow computation.

Suppose we have a continuous image where E(x,y,t) refers to the gray-level of (x,y) at time t representing the dynamic image as a function of position and time permits it to be expressed as a Taylor series:
E(x+uδt,y+vδt,t+δt)=E(x,y,t)+Exδx+Eyδy+Etδt+O(δt^2)

Where Ex, Ey, Et denote the partial derivatives of E. The u(x,y) and v(x,y) are the components of optical flow. We can assume that the immediate neighborhood of (x,y) is translated some small distance (δx,δy) during the interval δt; that is, we can find δx, δy, δt such that

E(x+uδt,y+vδt,t+δt)=E(x,y,t)+Exδx+Eyδy+Etδt

The above equation is also known as brightness conservation equation. If δx, δy, δt are very small, the higher order terms in the equation vanishes. Dividing by δt and taking the limit δt→0, leads to the following expression.

Therefore, the brightness constraint equation is given by,

Ex u + Ey v + Et = 0

Assuming the global smoothness of the brightness changes in the images, one can model the motion field applying the higher order derivatives of the data conservation equation. Iterative solutions of these two or more equations or certain regression methods applied on the relevant set of equations yield the components of the velocity vector field. In the present work we have computed the following differences and are used suitably.

3. Algorithm

Line - by- line algorithm is presented below.

1. Selection of reference image in an image sequence: Image with no traffic in the identified sequence is considered as the reference image.

2. Dynamic selection of reference/background images. Compute normalized difference, \( d_{nor}(i,j) = \frac{1}{N \times M} \sum_{l} \{ d_l(i,j) - d_{l+1}(i,j) \} \)

3. Selection of Background Image for computation of flux:
   - If \( Er = \sum |d_l(i,j) - d_{l+1}(i,j)| \leq \hat{\alpha}_1 \), then skip one image \( I = I + 1 \)
   - If \( Er = \sum |d_l(i,j) - d_{l+1}(i,j)| > \hat{\alpha}_1 \) and \( \leq \hat{\alpha}_2 \), then skip three images \( I = I + 2 \)
   - If \( Er = \sum |d_l(i,j) - d_{l+1}(i,j)| > \hat{\alpha}_2 \) and \( \leq \hat{\alpha}_3 \), then skip five image \( I = I + 3 \)

Where \( \hat{\alpha}_1 < \hat{\alpha}_2 < \hat{\alpha}_3 \) Further this set can be re-asserted.

4. Selected image is passed through spatial low pass filter.

5. Find difference image between incoming video frame and background image.

6. using multiple thresholds create binary image

7. Compute the normalized average brightness of the segmented region and compute flux as a percentage with reference background.

8. Vehicle Count :

\[ \text{Slope} = \phi(n+1) - \phi(n) \]
\[ \phi(n) = \text{present flux density value for specific frame period} \]
\[ \phi(n) = \text{previous flux density value} \]

\[ \text{Sign} = \text{sign}(x) = y \]

\[ \{ 1 \text{ if } y_{n+1} = 1 \text{ and } y_n = -1 \] \[
\{ \text{Vehicle} = 0 \text{ if } y_{n+1} = -1 \text{ and } y_n = 1 \]

1744
Detection, Counting and Classification of Moving Objects by Using Real Time Traffic Flux through Differential and Rule Based Analysis

\[ \begin{cases} 0 \text{ if } y_{n+1} = 1 \text{ and } y_n = 1 \\ 0 \text{ if } y_{n+1} = -1 \text{ and } y_n = -1 \end{cases} \]

9. Object classification:

\[ \Phi \rightarrow \text{percentage flux density} \]

Classification:
- P \rightarrow \text{Pedestrian}
- T \rightarrow \text{Two wheeler}
- F \rightarrow \text{Four wheeler}
- H \rightarrow \text{Heavy vehicles}

Case (i)

\[ \Phi_{p_i} \leq \Phi \leq \Phi_{p_{i+1}} \]

- Pedcount = Pedcount
- Pcount = Pcount + 1
- Pecount = Pcount

10. Step nine is repeated for other type of vehicles with different slopes, rate of change of slopes and end flux values.

4. Implementation

Following general considerations and assumptions are made in order to implement the computer vision based system for traffic flux estimation.

- Camera is positioned at a fixed location with predetermined focus. This is in order to eliminate ego-motion problem.
- Video sequences are taken from the top to minimize occlusion also perspective oblique view to validate the result.
- Fixed number of frames with fixed camera resolution.
- Both color and black and white videos are used.
- Monocular video is the input for processing.

The block diagram of the implemented system is shown in Fig. 1. The video input is given to the system with an input video rate at 30 frames per second. Frames are separated using frame grabber software. Further frames are converted to BMP format for the purpose of analysis. Pre processing follows which encompasses the low pass filters and spatial threshold operations. Cumulative and Normalized difference between frames are computed and used to establish the traffic flux.

At first we will consider a continuous family of images on some time period and derive expressions for traffic flux in terms of spatial and temporal derivatives of this continuous image sequence with \( \Phi \rightarrow \mathbb{R} \rightarrow \mathbb{R} \) \( \Phi \rightarrow \mathbb{R} \rightarrow \mathbb{R} \rightarrow E(x, y) ; \) Here \( \mathbb{R} \) is the image domain. We will always denote the sequence parameters by \( r \) and \( s \) respectively, where as \( x \) and \( y \) respectively stands for the spatial co-ordinates. We assume \( \Phi \) to be smooth in time and space.

![Figure 1](image)

5. Discussion on Results

In order to test the above-developed algorithm, several sets of natural image sequences are used. Real image sequences, recorded in MPEG2 format have been used with camera, in fixed position to capture the aerial view of the road. Different natural traffic videos are taken in...
situations where obstacles are found in the line of view, vehicle shadows, building shadows in the path and oblique view of the traffic. The first set of images is taken in order to establish the reference images under different illumination condition from morning to evening. Four such reference frames have been identified under supervision.

In the present work, a platform has been created so that complete automation of dynamic and intelligent traffic control devoid of human intervention. The incorporated dynamic selection of background images supports the selection of required frames [12]. Subtraction of images in the raw data form helps to generate boundary between moving objects in the scene with the background. After subtraction the image is negated. Four levels of threshold is implemented to mark the difference between two images and assigned with different colors in the output image [11]. From the results obtained it is evident that the implemented algorithm is separating object from the background pixels, detection and classification of objects in almost all the cases. Also, it works satisfactorily when there is a change in illumination, obstacles found in the image path and building shadows.

Confidence Measure: One of our major investigations has been the identifying confidence measure to establish the validity of the results. This provides means of determining the reliability of the computed object classification. Traffic flux computed is verified using tessellation method. Vehicle classification is verified through DFT technique and visual analysis.

Table 1: Results of Object Classification in a Typical Case

<table>
<thead>
<tr>
<th>Objects/vehicles</th>
<th>count</th>
<th>Error positive</th>
<th>Error negative</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedestrian</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Two wheeler</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>88.00%</td>
</tr>
<tr>
<td>Four wheeler</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Heavy vehicles</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>97.00%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

The work carried out has produced very good and consistent results. The small deviations with different video streams taken under different situations are minimum and do not really hampers traffic analysis. Following are the limitations observed by the authors at this juncture and further finds the scope to continue the work.

1. The shadows of the vehicles and buildings in the field of view are causing subtle variation in the traffic flux and vehicle count computation.
2. If the color of the vehicle and the color of the road (Reference image) are same it may lead to marginally varied traffic (approximately 5-6%) flux and as well as vehicle count computation.
3. With the present computational facility maximum size of the color image that can be handled in real time is 200 pixels by 200 pixels.
4. Variation in illumination condition causes subtle variation in estimated traffic flux and vehicle count.
5. Depletion in accuracy to detect and classify the objects when there are multiple objects present having simultaneous entry and exit because of resulted effect on percentage flux.

7. Presentation and Display of Results
The first level result shows the computation of traffic flux under six different natural conditions. The results are shown in figure 2. It consists of traffic flux graph, sample frames and table consisting of traffic flux values and the difference of flux. Peaks and valleys are indicated as +1 and -1. Further the change of sign increments the vehicle count. All images are of 200*200 sizes. Here result shown for only one pair of images in the sequence. Such computation is being done for all the images in the video input to the system. The result presented ensures the correctness of the computation as it is more near to the visual estimation value, which uses tessellation method. In figure 3, second set of result shows the vehicle classification. Four different types of object/vehicles have been presented along with their flux graph.

References

Author's Biography

Prof. H.S. Mohana

Obtained B.E Degree in Electrical and Electronics Engineering from University of Mysore during 1986. Since then serving technical education field in various capacities. Obtained M.E from University of Roorkee, presently IIT ROORKEE with the specialization in Measurement and Instrumentation. Worked as chairmen and Member of Board of Examiner and Board of studies with several universities which include, University of Mysore, Kuvempu University and VTU. Presented research findings in 12 National Conferences and in 4 International conferences held across the world. Recognized as AICTE expert committee member in the inspection and reporting continuation of affiliation and Increase in intake of the Engineering Colleges. Completed, one AICTE/MHRD-TAPTECH project, and one AICTE/MHRD- Research project successfully. Coordinated TWO ISTE Sponsored STTP for the technical college teachers.

Dr. Aswatha Kumar. M has obtained B.E Degree from University of Mysore in Electronics and communication Engineering. Since then serving technical education field in various capacities. Obtained M.E from IISc Bangalore in ECE Department. Carried out research under the guidance of Professor B. N. Chattergi at IIT Kharagpur and obtained PhD. Worked as chairmen and Member of Board of Examiner and Board of studies with several universities which includes, University of Mysore, Kuvempu University and VTU. Chaired technical sessions at many National and International conferences. Presented research findings in 35 National Conferences and International conferences held across the world. Organized, many National and International conferences. Completed, many AICTE/MHRD-TAPTECH project, and one AICTE/MHRD- Research project successfully. Working as referee for reputed National and International Journals. Worked as technical consultant for many industries and banks. Presently, five research scholars pursuing research work.
Result Images and Graphs

For two similar IMAGES resultant subtracted image is shown

Figure 2 : Output Graph and Vehicle Count
Result Images and Graphs for Classification

Pedestrian

Two Wheelers
Detection, Counting and Classification of Moving Objects by Using Real Time Traffic Flux through Differential and Rule Based Analysis

Four wheelers

Heavy Vehicles

Figure 3: Output Graph and Object Classification

1751
A Hierarchical Automatic Language Identification System for Indian Languages Using Acoustic Features

S. Jothilakshmi1 V. Ramalingam2 S. Palanivel3

1, 2 & 3Department of Computer Science and Engineering, Annamalai University, Annamalainagar-608 002, India.
Email: jothi.sekar@gmail.com, aucsevr@yahoo.com, spal.yughu@yahoo.com

ABSTRACT

Automatic spoken language identification (LID) is the task of identifying the language from a short utterance of the speech signal uttered by an unknown speaker. This paper describes a novel two level identification system for Indian languages using acoustic features. In the first level, the system identifies the family of the spoken language, and then it is fed to the second level which aims at identifying the particular language in the corresponding family. The proposed system has been modelled using Hidden Markov Model (HMM) and utilizes the acoustic features namely Mel frequency cepstral coefficients (MFCC) and Shifted delta cepstrum (SDC). A new database has been created for 11 Indian languages. The proposed system achieves a high accuracy of 62.36% for MFCC features and 71.2% for SDC features.

Keywords: Language identification, Indian languages, Hidden Markov model, Mel frequency cepstral coefficients (MFCC), Shifted delta cepstrum (SDC).

1. INTRODUCTION

The automatic language identification (LID) [1], [2] is a process by which the language spoken in a particular speech utterance is identified. It is an important technology in many applications, such as spoken language translation [3], multi lingual speech recognition [4], and spoken document retrieval [5].

Humans are the best LID systems in the world today. Just by hearing one or two seconds of speech of a familiar language, they can easily identify the language. The performance of any LID system depends on the amount of information and the reliability of information extracted from the speech signal and how efficiently it is incorporated into the system [1].

Existing spoken language identification systems can be broadly classified into two groups namely, explicit and implicit systems, The LID systems that require speech recognizers of one or several language, in other words, the systems that require a segmented and labelled speech corpus are termed as explicit LID systems. The language identification systems which do not require phone recognizers (or rather segmented and labelled speech data) are termed here as implicit LID systems. In other words, these systems require only the raw speech data along with the true identity of the language spoken [1], [6].

In general, LID features fall into five groups according to their level of knowledge abstraction [7]. Lower level features, such as spectral feature, are easier to obtain but volatile because speech variations such as speaker to channel variations are present. Higher level features, such as lexical /syntactic features, rely on large vocabulary speech recognizer, which is language and domain dependant. They are therefore difficult to generalize across languages and domains. Phonotactic features become a trade-off between computational complexity and performance. It is generally agreed that phonotactics i.e. the rules governing the sequences of admissible
phones/phonemes, carry more language discriminative information than the phonemes themselves. This work focuses on acoustic only LID system for which Hidden Markov Modelling is the state of the art classifier.

In this paper, a hierarchical language identification system has been proposed for Indian languages. The languages of India belong to four major linguistic families namely Indo Aryan, Dravidian, Austro-Asiatic and Tibeto-Burman [8]. The largest of these in terms of speakers is Indo Aryan which is spoken by 75.278% of the people. The second largest is the Dravidian family which is spoken by 22.5% of the people whereas Austro Asiatic is spoken by 1.132% of the people and Tibeto-Burman by 0.965%. So, nearly 98% of the people in India are speaking languages from Aryan family and Dravidian family. Hence the proposed two level systems are designed to identify the languages of Aryan and Dravidian family. In the first level, the system identifies the family of the spoken language, and then it is fed to the second level which aims at identifying the particular language in the corresponding family.

The aim of this work is to design a less complex system for Indian language identification, so only the acoustic features are utilized in the system. The most widely used features for LID are Mel frequency cepstral coefficients (MFCC). Traditionally, language and speaker identification tasks use feature vectors containing cepstra and delta and acceleration cepstra. Recently, however, the shifted delta cepstrum (SDC) has been found to exhibit superior performance to the delta and acceleration cepstra in a number of language identification studies [9] due to its ability to incorporate additional temporal information, spanning multiple frames, into the feature vector. The performances of MFCC and SDC features are compared in this paper.

The paper is organized as follows: section 2 briefly reviews the feature extraction used for LID. In section 3, we briefly review Hidden Markov model. Our method is presented in detail in section 4. Experimental settings and results are described in section 5. Finally the conclusion is given in section 6.

2. FEATURE EXTRACTION FOR LANGUAGE IDENTIFICATION

Speech signals need to be parameterized prior to identification process. Parameterization consists of the extraction of a set of features from the speech waveform, which may present two main characteristics: they must provide a reasonable and compact representation of the speech signal and they must have adequate discrimination capabilities for discriminating between sounds.

2.1 Mel Frequency Cepstral Coefficients (MFCC)

Mel frequency cepstral coefficients (MFCC) have proved to be one of the most successful feature representations in speech related recognition tasks [10]. The mel-cepstrum exploits auditory principles, as well as the decorrelating property of the cepstrum. The computation of MFCC is shown in Fig.2.1 and described as follows.

![Figure 2.1: Extraction of MFCC from Speech Signal](image)

**Preemphasis**

The digitized speech signal $s(n)$ is put through a low order digital system (typically a first-order FIR filter), to spectrally flatten the signal and to make it less susceptible to finite precision effects later in the signal processing. The output of the preemphasis network, $\hat{s}(n)$ is related to the input $s(n)$, by the difference equation
\( \hat{s}(n) = s(n) - \alpha s(n - 1) \)

The most common value for \( \alpha \) is around 0.95.

**Frame Blocking**

Speech analysis usually assumes that the signal properties change relatively slowly with time. This allows examination of a short time window of speech to extract parameters presumed to remain fixed for the duration of the window. Thus to model dynamic parameters, the signal must be divided into successive windows or analysis frames, so that the parameters can be calculated often enough to follow the relevant changes. In this step the preemphasized speech signal, \( \hat{s}(n) \), is blocked into frames of \( N \) samples, with adjacent frames being separated by \( M \) samples. If we denote the \( l \)th frame speech by \( x_l(n) \), and there are \( L \) frames within the entire speech signal, then

\[
1, ..., 0, 1, ..., 0, 1, ..., 0, 1, ..., 0, 1, ..., 0, 1, ..., 0, \hat{s}(n), 1, ..., 0 = + = - = \sum_{n=M}^{L} n \sum_{n=0}^{N} x_l(n)
\]

**Windowing**

The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of the frame. The window must be selected to taper the signal to zero at the beginning and end of each frame. If we define the window as \( w(n) \), \( 0 \leq n \leq N - 1 \) then the result of windowing the signal is

\[
w(n) = 0.54 - 0.46 \cos \left( \frac{2\pi n}{N - 1} \right) \quad 0 \leq n \leq N - 1
\]

**Computing Spectral Coefficients**

The spectral coefficients of the windowed frames are computed using Fast Fourier Transform, as follows:

\[
\sum_{n=0}^{N-1} x_l(n) e^{-j2\pi nk/N} = X_k
\]

**Computing mel Spectral Coefficients**

The spectral coefficients of each frame are then weighted by a series of filter frequency response whose center frequencies and bandwidths roughly match those of the auditory critical band filters. These filters follow the mel scale whereby band edges and center frequencies of the filters are linear for low frequency and logarithmically increase with increasing frequency as shown in Fig. 2.2. These are called as mel-scale filters and collectively a mel-scale filter bank [11]. As can be seen, the filters used are triangular and they are equally spaced along the mel scale which is defined by

\[
\text{Mel}(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)
\]

**Figure 2.2 : Mel- Scale Filters**

Each short term Fourier transform (STFT) magnitude coefficient is multiplied by the corresponding filter gain and the results are accumulated.

**Computing MFCC**

The discrete cosine transform (DCT) is applied to the log of the mel spectral coefficients to obtain the MFCC as follows:

\[
x(m) = \sqrt{\frac{2}{M}} \sum_{n=0}^{M-1} \hat{X}(n) \cos \left( \frac{(2m+1) \pi n}{2M} \right) \quad m = 1, ..., M
\]
Where $M$ is the number of filters in the filter bank, finally, cepstral mean subtraction is performed to reduce the channel effects.

2.2 Shifted Delta Cepstrum (SDC)
The shifted delta cepstral features have been introduced to improve the LID performance with respect to the classical cepstral and delta cepstral features [12].

![Figure 2.3: Calculation of the Shifted Delta Feature Vectors](image)

The SDC coefficients are computed, for a cepstral frame at time $t$, according to:

$$D_i P_t c_n = \Delta^1 \Delta^0 c_n - \Delta^0 c_n$$

Where $n$ is the $n^{th}$ cepstral coefficients, $D$ is the lag of the deltas, $P$ is the distance between successive delta computations, and $i$ is the SDC block number. The final feature vector is obtained by concatenation of $k$ blocks of $N$ parameters.

The computation of the Shifted Delta feature vectors is a relatively simple procedure. The process is as follows: The MFCC feature vectors are first computed as described above. Then, the acoustic feature vectors spaced $D$ sample frames apart are first differenced. Then $k$ differenced feature vector frames, spaced $P$ frames apart, are then stacked to form a new feature vector. Fig.2.3 gives a graphical depiction of this process.

3. Hidden Markov Model

Hidden Markov model [11], [13] is used in the problem of making a sequence of decisions on temporal basis. It is a statistical model and a variant of finite state machine. In Markov model the states are directly accessible to the observer. But in HMM the states are not directly accessible to the observer only the variables influenced by the states are accessible to the observer.

3.1 Notations used in HMM

- $w \rightarrow$ Hidden State
- $v \rightarrow$ Visible state
- $a_{ij} \rightarrow$ Transition probability to make transition from $i^{th}$ state at $t$ to $j^{th}$ state at $(t+1)$
- $b_{jk} \rightarrow$ Emission probability to emit $k^{th}$ visible state at $j^{th}$ hidden state.

$N \rightarrow$ Number of hidden states (Guess this number)
$M \rightarrow$ Number of visible states (obtained from the training set)

3.2 Design Issues

The HMM will be useful in real world applications, if three basic problems of HMM are solved. These problems are the following.

1. Learning problem.
2. Evaluation problem.
3. Decoding problem

**Learning Problem**

Given the values of $N, M$:

- The goal of learning is to determine model parameters-{$a_{ij}, b_{jk}$} from the training samples.
- Forward – backward algorithm, also known as Baum Welch algorithm is used for learning problem.
Forward algorithm will generate $\alpha$ values. By using backward algorithm we should find $\beta$ value.

Let the model is in state $w(t)$ by generating part of the given visible sequence, $\alpha$ is nothing but the probability taken so far to come to the current state from the initial state. We express $\alpha_i(t)$ as

$$\alpha_i(t) = \begin{cases} 0 & ;w_i(t) \neq 1 \text{ for final state and } t = T \\ 1 & ;w_i(t) = 1 \text{ for final state and } t = T \\ \sum_{j} \beta_j(t+1) a_{ij} b_{jk} v(t+1) & ;\text{Otherwise} \end{cases}$$

Evaluation process started initially by randomly selecting the value of $a_{ij}$ and $b_{jk}$ (such that the summation of each row of $a_{ij}$ and $b_{jk}$ is equal to 1). Then re-estimation of $a_{ij}$ and $b_{jk}$ will be done to achieve the true values of $a_{ij}$ and $b_{jk}$. For the same training data again $\theta$ is calculated by using re estimated $\{a_{ij}\} \text{ and } \{b_{jk}\}$. This re-estimation for the same training data will be done repeatedly until the value of $\{a_{ij}\}$ and $\{b_{jk}\}$ is constant for subsequent iterations or negligible change in the estimated values of the parameters on subsequent iterations. Now the values of $\{a_{ij}\} \text{ and } \{b_{jk}\}$ are the true values. So it can be applied to test data.

Evaluation Problem

The goal is to find the probability to generate a particular sequence of visible state $V^T$ by the model when the HMM parameters $\theta$ is given. $\theta = \{a_{ij}, b_{jk}\}$. The probability of each possible sequence of hidden states to produce $V^T$ is calculated and then the probabilities are added up. So

$$P \left( V^T \mid W^T \right) = \sum_{i=1}^{N} P \left( V^T, W^T \mid \theta \right)$$

But this type of calculation is much complex. It will take $O(NTT)$ calculation. A computationally simpler recursive algorithm for the same goal is the forward algorithm.

Decoding Problem

The decoding problem is to find the most probable sequence of hidden states for the given sequence of visible states $V^T$. For decoding Viterbi algorithm is used. The decoding algorithm finds at each time step $t$, the state that has the highest probability $\alpha_j(t)$. The full path is the sequence of hidden states to generate the given visible state sequence optimally.

4. Language Identification System

The acoustic systems are an interesting compromise between complexity and performance. We have implemented a simple acoustic system for Indian languages using MFCC, SDC coefficients and Hidden Markov model. An acoustic language identification system based on Hidden Markov Model (HMM) works in two phases, a learning procedure to create the models, and a testing procedure.

To identify $N$ number of languages, $N$ numbers of HMMs are to be modeled because each language is to be modeled by a distinct HMM. For each language, a training set of $K$ speech segments spoken by many talkers. For each language many observation sequences $(V^T)$ will be there. Each HMM will be trained by using the observation sequences of the corresponding language by doing
A Hierarchical Automatic Language Identification System for Indian Languages Using Acoustic Features

learning process. The training will be stopped after obtaining optimal values for \( \alpha \) and \( \beta \). Likewise all \( N \) HMMs will be trained.

During testing phase, each unknown speech segment, the language of which is to be identified is applied to the system. The observation sequences \( V^T \) are obtained and applied to all HMMs (from HMM \( _0 \) to HMM \( _N \)). Each HMM will compute the probability \( P \left( \frac{V^T}{\theta} \right) \) for the particular observation sequence \( V^T \) by using evaluation process. From all \( P \left( \frac{V^T}{\theta} \right) \) values the maximum value will be selected by using Viterbi algorithm. This is the unknown language, i.e. for the particular language generated by the corresponding HMM will be greater than other HMMs.

Even though many languages are in Aryan family, the languages spoken by large number of peoples are considered in this system. The languages spoken by less than 2% of the people of the country are not included.

5. Experiments And Results

5.1 The Database

All the experiments described in this paper were conducted on our own database. It comprises broadcast news shows in 11 languages. In Dravidian family, all four languages namely Tamil (Ta), Telugu (Te), Kannada (Ka) and Malayalam (Ma) languages are used. In Aryan family the major languages namely Hindi (Hi), Bengali (Be), Marathi (Mar), Gujarati (Gu), Oriya (Or), Kashmiri (Kas) and Punjabi (Pu) are selected. This database contains a total of 10h of broadcasts from Doordharasen television network because the network is available in all regional languages of India.

Train and test sets have been defined for each language. For each language, 30 speakers are selected as the training set, and the duration of each speaker is about 60 seconds. The testing set consists of 10 speakers and the duration of each speaker is 10 seconds.

5.2. Feature Extraction

The selected properties for the speech signals are a sampling rate of 8 kHz, 16 bit monophonic PCM format. We used a frame rate of 125 frames/s, where each frame is 16ms in duration with an overlap of 50% between adjacent frames. All the training and test data are pre-processed to remove silence from the speech signals.

The feature vectors used consist of 13 Mel frequency cepstral coefficients (MFCC). Finally, the delta and acceleration coefficients are appended to the features. So for each frame, a 39 dimensional feature vector is calculated.

Figure 4.1: Overview Of The Proposed Acoustic Language Identification System

The proposed system for identifying Indian languages is a two level system as shown in Fig. 4.1. In the first level, it will identify whether the language belongs to Dravidian family or Aryan family. Then in the second level it will identify the corresponding language.

In Dravidian family, all languages namely Tamil, Telugu, Kannada and Malayalam are considered in this system.
The configuration 7-1-3-7 for N-D-P-k has been used to extract SDC feature vector. For each frame, with 7 direct MFCC coefficients 49 SDC coefficients are appended, so totally 56 coefficients are used.

5.3 Hidden Markov Model Classifier
The first level of the system uses two HMMs and the second level uses four HMMs for Dravidian family and seven HMMs for Aryan family. All HMMs are initialized with five states and two Gaussian mixtures/state.

5.4 Results
Investigations were conducted to compare the performance of the system with MFCC with delta and acceleration coefficients and SDC individually. The performance of the system is given in Table 1. The results of experiments indicate that the proposed system is able to help in distinguishing between languages with greater accuracy. The average performance is affected by the poor performance for the languages Kannada and Oriya.

5.5 Discussion
The major challenge in Indian languages is the similar characteristics of the languages. So it is very difficult to distinguish one from the other and it is a challenging task to design a language identification system for these languages. In this system, we used continuous speech for both training and testing. The purpose of hierarchical system is to reduce the complexity of the system. Once the family is identified then it is enough to compare the test utterance within the languages of the family. But the drawback in multi stage system is each stage of the multi stage model exploits results from the previous stage, errors introduced by a stage certainly affects the accuracy of next stage. In this system also the second level results are affected by the first stage. Here we selected the features and all the parameters based on the best features and best parameter values used in the existing LID systems.

6. Conclusion
In this work a novel two level language identification system is proposed for Indian languages using acoustic features. The acoustic systems are an interesting compromise between complexity and performance. Investigations were conducted to compare the performance of the system with MFCC with delta and acceleration coefficients and SDC individually. The proposed system has been designed to identify 11 major Indian languages. We created a new database to investigate the performance of this system. The system with SDC performs better than the system with MFCC features.

In future the research will be conducted in the direction to improve the performance of the system. This can be achieved by combining the prosodic features with the acoustic features, by using other modeling techniques and by improving the training and testing data sets. As the Indian languages are similar in characteristics, designing a less complex system with the best performance is a challenging task. This work is the first step in this direction.

<table>
<thead>
<tr>
<th>Language</th>
<th>MFCC</th>
<th>SDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ta</td>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>Te</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Ma</td>
<td>60</td>
<td>75</td>
</tr>
<tr>
<td>Ka</td>
<td>35</td>
<td>60</td>
</tr>
<tr>
<td>Gu</td>
<td>60</td>
<td>70</td>
</tr>
<tr>
<td>Mar</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>Pu</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>Hi</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>Be</td>
<td>76</td>
<td>88</td>
</tr>
<tr>
<td>Kas</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>Or</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Average</td>
<td>62.36</td>
<td>71.2</td>
</tr>
</tbody>
</table>

Table 1 : Language wise Performance in %
A Hierarchical Automatic Language Identification System for Indian Languages Using Acoustic Features

REFERENCES


Author’s Biography

S. Jothilakshmi received the B.E degree in Electronics and Communication Engineering from Govt. College of Engineering, Salem in 1994. She received the M.E. degree in Computer Science and Engineering from Annamalai University in the year 2005. She has been with Annamalai University, since 1999. She is pursuing her Ph.D degree in Computer Science and Engineering at Annamalai University. She published 5 papers in international conferences and journals. Her research interest includes speech processing, image and video processing, and pattern classification.

Dr. V. Ramalingam received the M.Sc degree in Statistics from Annamalai University in 1980. He received the M.Tech degree in Computer Applications from Indian Institute of Technology Delhi in the year 1995. He has been with Annamalai University, since 1982. He completed his Ph.D degree in Computer Science and Engineering at Annamalai University in 2006. He published 27 papers in international conferences and journals. His research interest includes image and video processing, natural language processing and neural networks.
Dr. S. Palanivel received the B.E (Hons) degree in Computer Science and Engineering from Mookambigai College of Engineering in 1989. He received the M.E degree in Computer Science and Engineering from Government College of Technology in the year 1994. He has been with Annamalai University, since 1994. He completed his Ph.D degree in Computer Science and Engineering at Indian Institute of Technology Madras under quality improvement programme (QIP) sponsored by Annamalai University, in the year 2005. He published 27 papers in international conferences and journals. His research interest includes speech processing, image and video processing, pattern classification and neural networks.
ABSTRACT

PID controllers are widely used in industrial plants because it is simple and robust. Industrial processes are subjected to variation in parameters and parameter perturbations, which when significant makes the system unstable. The aim of this paper is to design a controller of a various plant by selection of PID parameters using soft computing techniques. Z-N methods whose performance have been compared and analyzed with the intelligent tuning techniques like Genetic algorithm, Evolutionary programming and particle swarm optimization. Soft computing methods have proved their excellence in giving better results by improving the steady state characteristics and performance indices.

Keywords: Genetic Algorithm, Evolutionary Programming, Particle Swarm Optimization And Soft Computing.

1. INTRODUCTION

Conventional proportional integral derivative controller is widely used in much industrial application due to its simplicity in structure and ease to design [1]. However it is difficult to achieve the desired control performance. Tuning is important parameter for the best performance of PID controllers. PID controllers can be tuned in a variety of ways including hand tuning Ziegler Nichols tuning[14], Cohen-coon tuning and Z-N step response, but these have their own limitations [3]. Soft computing techniques like GA, PSO and EP methods have proved their excellence in giving better results by improving the steady state characteristics and performance indices.

1.1. Proportional Integral Derivative Controller

The PID controller calculation involves three separate parameters proportional integral and derivative values. The proportional value determines the reaction of the current error, the integral value determines the reaction based on the sum of recent errors, and derivative value determines the reaction based on the rate at which the error has been changing the weighted sum of these three actions is used to adjust the process via the final control element[13].

The block diagram of a control system with unity feedback employing Soft computing PID control action in shown in figure 1 [7].

\[ Y(t) = [k_p(e(t)) + k_d \frac{d(e)}{d(t)} + k_i \int_0^t e(t)d(t)] \]  

(1)

Figure 1: Block Diagram of Intelligent PID Controller

1 & 2 Tamilnadu news print and Paper Ltd, Tamilnadu, India.
3 Karpagam University, India. Email: nagarajice@gmail.com
2. REASON FOR SELECTING SOFT COMPUTING TECHNIQUES

Model Type

Many methods can be used only when the process model is of a certain type, for example a first order plus dead time model (FOPDT). Model reduction is necessary if the original model is too complicated. [6]

Design Criteria: These methods aim to optimize some design criteria that characterize the properties of the closed-loop system. Such criteria are, for example, gain and phase margins, closed-loop bandwidth, and different cost functions for step and load changes. [6]

Approximations: Some approximations are often applied in order to keep the tuning rules simple. [6]

The purpose of this project is to investigate an optimal controller design using the Evolutionary programming, Genetic algorithm, Particle swarm optimization techniques. In this project, a new PID tuning algorithm is proposed by the EP, GA, and PSO techniques to improve the performance of the PID controller.

The ultimate gain and the ultimate period were determined from a simple continuous cycle experiment. The new tuning algorithm for the PID controller has the initial value of parameter $K_p$, $T_i$, $T_d$ by the Ziegler-Nichols formula that used the ultimate gain and ultimate period from a continuous cycle experiment and we compute the error of plant response corresponding to the initial value of parameter.

The new proportional gain ($K_p$), the integral time ($T_i$), and derivative time ($T_d$) were determined from EP, GA, and PSO. This soft computing techniques for a PID controller considerably reduced the overshoot and rise time as compared to any other PID controller tuning algorithms, such as Ziegler-Nichols tuning method and continuous cycling method.

2.1 Genetic Algorithm

Genetic Algorithms (GA.s) are a stochastic global search method that mimics the process of natural evolution. It is one of the methods used for optimization. John Holland formally introduced this method in the United States in the 1970 at the University of Michigan. The continuing performance improvement of computational systems has made them attractive for some types of optimization. The genetic algorithm starts with no knowledge of the correct solution and depends entirely on responses from its environment and evolution operators such as reproduction, crossover and mutation to arrive at the best solution [1]. By starting at several independent points and searching in parallel, the algorithm avoids local minima and converging to sub optimal solutions.

2.1.1 Objective Function of the Genetic Algorithm

This is the most challenging part of creating a genetic algorithm is writing the objective functions. In this project, the objective function is required to evaluate the best PID controller for the system. An objective function could be created to find a PID controller that gives the smallest overshoot, fastest rise time or quickest settling time. However in order to combine all of these objectives it was decided to design an objective function that will minimize the performance indices of the controlled system instead. Each chromosome in the population is passed into the objective function one at a time. The chromosome is then evaluated and assigned a number to represent its fitness, the bigger its number the better its fitness [3]. The genetic algorithm uses the chromosomes fitness value to create a new population consisting of the fittest members. Each chromosome consists of three separate strings constituting a P, I and D term, as defined by the 3-row bounds declaration when creating the population [3]. When the chromosome enters the evaluation function, it is split up into its three Terms. The
newly formed PID controller is placed in a unity feedback loop with the system transfer function. This will result in a reduce of the compilation time of the program. The system transfer function is defined in another file and imported as a global variable. The controlled system is then given a step input and the error is assessed using an error performance criterion such as Integral square error or in short ISE. The chromosome is assigned an overall fitness value according to the magnitude of the error, the smaller the error the larger the fitness value.

2.1.1 Evolutionary Programming

There are two important ways in which EP differs from GAs.

First, there is no constraint on the representation. The typical GA approach involves encoding the problem solutions as a string of representative tokens, the genome. In EP, the representation follows from the problem. A neural network can be represented in the same manner as it is implemented, for example, because the mutation operation does not demand a linear encoding [6].

Second, the mutation operation simply changes aspects of the solution according to a statistical distribution which weights minor variations in the behavior of the offspring as highly probable and substantial variations as increasingly unlikely.

The steps involved in creating and implementing evolutionary programming are as follows:

- Generate an initial, random population of individuals for a fixed size (according to conventional methods \( K_p, T_i, T_d \) ranges declared).
- Evaluate their fitness (to minimize integral square error).
- Select the fittest members of the population.
- Execute mutation operation with low probability.
- Select the best chromosome using competition and selection.
- If the termination criteria reached (fitness function) then the process ends. If the termination criteria not reached search for another best chromosome.

2.1.1.1 Particle Swarm Optimization

PSO is one of the optimization techniques and kind of evolutionary computation technique [15]. The technique is derived from research on swarm such as bird flocking and fish schooling. In the PSO algorithm, instead of using evolutionary operators such as mutation and crossover to manipulate algorithms, for a \( d \)-variable optimization Problem, a flock of particles are put into the \( d \)-dimensional Search space with randomly chosen velocities and positions knowing their best values.

So far (p best) and the position in the \( d \)-dimensional space \( V_i \) [7]. The velocity of each particle, adjusted accordingly to its own flying experience and the other particles flying experience [7].

For example, the \( i \)th particle is represented, as \( x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,d}) \)

In the \( d \)-dimensional space. The best previous position of the \( i \)th particle is recorded as,

\[
P_{\text{best}_i, d} = (P_{\text{best}_{i,1}}, P_{\text{best}_{i,2}}, \ldots, P_{\text{best}_{i,d}}) \quad (2)
\]

The index of best particle among all of the particles in the group in \( g \) best \( d \) .The velocity for particle \( i \) is represented as

\[
V_i = (V_{i,1}, V_{i,2}, \ldots, V_{i,d}) \quad (3)
\]

The modified velocity and position of each particle can be calculated using the current velocity and distance from \( P_{\text{best}_i, d} \) to \( g_{\text{best}_d} \) as shown in the following formulas.
\[ V_{i,m}^{(t+1)} = W V_{i,m}^{(t)} + c_1 \times \text{rand}(t) \times (P_{\text{best}_{i,m}} - x_{i,m}^{(t)}) + c_2 \times \text{rand}(t) \times (g_{\text{best}_{m}} - x_{i,m}^{(t)}) \] — (4)

\[ x_{i,m}^{(t+1)} = x_{i,m}^{(t)} + V_{i,m}^{(t+1)} \]

\( i = 1, 2, \ldots, n \)
\( m = 1, 2, \ldots, d \)

Where

\( n \) = Number of particles in the group

\( d \) = dimension

\( t \) = Pointer of iterations (generations)

\( V_{i,m}^{(t)} \) = Velocity of particle I at iteration t

\( W \) = Inertia weight factor

\( c_1, c_2 \) = Acceleration constant

\( \text{rand}(d) \) = Random number between 0 and 1

\( x_{i,m}^{(t)} \) = Current position of particle i at iterations

\( P_{\text{best}_{i}} \) = Best previous position of the ith particle

\( g_{\text{best}_{m}} \) = Best particle among all the particles in the population

3. RESULTS AND DISCUSSIONS

In order to cover typical kinds of common industrial processes have been taken

Model-A

Model-B

Model-C

Model-D

Figure 3 : Flowchart of GA

Figure 2 : Flowchart of PSO

0.0147

0.0000207s + 0.000437

0.000077s^2 + 0.0539s^2 + 1.441s

15 \text{ s}^{-1}

Non optimum solution

Optimum Solution

0.9372s^3 + 2.656s^3 + 75.87s^2 + 112.1s

0.9372s^3 + 2.656s^3 + 75.87s^2 + 112.1s
3.1 Implementation of Intelligent PID Controller Tuning

The Ziegler-Nichols tuning method using root locus and continuous cycling method were used to evaluate the PID gains for the system, using the “rlocfind” command in matlab, the cross over point and gain of the system were found respectively.

In this paper a time domain criterion is used for evaluating the PID controller. A set of good control parameters P, I, and D can yield a good step response that will result in performance criteria minimization in the time domain.

These performance criteria in the time domain include the over shoot rise time and setting time. To control the plant model the following PSO, EP and GA parameters are used to verify the performance of the PID controller:

Performance characteristics of process model. A to D were indicated and compared with the intelligent tuning methods as shown in the figure 4 to figure 7 and values are tabulated in table II to table V.

Conventional methods of controller tuning lead to a large settling time, overshoot, rise time and steady state error of the controlled system. Hence Soft computing techniques is introduces into the control loop.

GA, EP and PSO based tuning methods have proved their excellence in giving better results by improving the steady state characteristics and performance indices.

Table 1: PSO, GA and EP Parameters

<table>
<thead>
<tr>
<th>PSO Parameters</th>
<th>GA Parameter</th>
<th>EP Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size: 100</td>
<td>Population size: 100</td>
<td>Population size: 100</td>
</tr>
<tr>
<td>Wmax = 0.6  Mutation rate: 0.1  Normal distribution</td>
<td>Cross</td>
<td>Mutation rate: 0.01</td>
</tr>
<tr>
<td>Iteration: 100</td>
<td>Fitnessfunction: ISE</td>
<td>Fitnessfunction: ISE</td>
</tr>
</tbody>
</table>

Figure 4: Comparison of All Methods for Model-A

Figure 5: Comparison of All Methods for Model-B

Figure 6: Comparison of All Methods For Model-C
Figure 7: Comparison of All Methods for Model-D

Table 2: Comparison Result of All Methods for Model -A

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Z-N</th>
<th>GA</th>
<th>EP</th>
<th>PSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Settling time (sec)</td>
<td>1.57</td>
<td>0.0098</td>
<td>0.0474</td>
<td>0.787</td>
</tr>
<tr>
<td>Rise Time (sec)</td>
<td>0.2</td>
<td>0.0055</td>
<td>0.0275</td>
<td>0.0663</td>
</tr>
<tr>
<td>Over shoot (%)</td>
<td>34</td>
<td>0.0042</td>
<td>0.528</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 3: Comparison Result of All Methods for Model -B

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Z-N</th>
<th>GA</th>
<th>EP</th>
<th>PSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Settling time (sec)</td>
<td>0.738</td>
<td>0.152</td>
<td>0.385</td>
<td>0.112</td>
</tr>
<tr>
<td>Rise Time (sec)</td>
<td>0.0375</td>
<td>0.063</td>
<td>0.015</td>
<td>0.001</td>
</tr>
<tr>
<td>Over shoot (%)</td>
<td>54.6</td>
<td>0.1</td>
<td>36</td>
<td>49.4</td>
</tr>
</tbody>
</table>

Table 4: Comparison Result of All Methods for Model -C

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Z-N</th>
<th>GA</th>
<th>EP</th>
<th>PSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Settling time (sec)</td>
<td>4.58</td>
<td>0.00315</td>
<td>0.134</td>
<td>0.0301</td>
</tr>
<tr>
<td>Rise Time (sec)</td>
<td>0.361</td>
<td>0.00257</td>
<td>0.0196</td>
<td>0.0246</td>
</tr>
<tr>
<td>Over shoot (%)</td>
<td>45</td>
<td>0.0365</td>
<td>26.6</td>
<td>0.224</td>
</tr>
</tbody>
</table>

Table 5: Comparison Result of All Methods for Model -D

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Z-N</th>
<th>GA</th>
<th>EP</th>
<th>PSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Settling time (sec)</td>
<td>20.4</td>
<td>0.023</td>
<td>0.43</td>
<td>0.0447</td>
</tr>
<tr>
<td>Rise Time (sec)</td>
<td>11.4</td>
<td>0.018</td>
<td>0.019</td>
<td>0.0365</td>
</tr>
<tr>
<td>Over shoot (%)</td>
<td>1</td>
<td>0.6</td>
<td>23</td>
<td>1</td>
</tr>
</tbody>
</table>

4. CONCLUSION

The GA, EP and PSO algorithm for PID controller tuning presented in this research offers several advantages. One can use a high-order process model in the tuning, and the errors resulting from model reduction are avoided. It is possible to consider several design criteria in a balanced and unified way. Approximations that are typical to classical tuning rules are not needed. Soft computing techniques are often criticized for two reasons: algorithms are computationally heavy and convergence to the optimal solution cannot be guaranteed. PID controller tuning is a small-scale problem and thus computational complexity is not really an issue here. It took only a couple of seconds to solve the problem. Compared to conventionally tuned system, GA, EP and PSO tuned system has good steady state response and performance indices.

REFERENCES


Optimum PID Controller Tuning Using Soft computing Methodologies for Industrial Process


Author's Biography

B.Nagaraj is a Assitant plant engineer at Tamilnadu nadu News print and paper ltd. He received his B.E. and M.E. degrees in Instrumentation and control engineering from Arulmigu Kalasalingam College of Engg in 2004 and 2006, respectively.He is pursing Ph.D degree in Karpagam university. He is a active member of instrument society of India, and Indian society for technical education. He has presented 31papers in National conferences and two papers in international conferences.

Energy and has contributed to various research innovations. He is co-director of NRL, Secretary-World Research scientist’s organization, Director-Futuristic Technology, MIT, Co-ordinator, Defense Research Board, US Govt. He holds 24 patents in various fields of research, 2 patents on concepts, 19 publications. He is the Honorary member of the Jack Welsh Science and technology, JSF, Asian federation of engineers, GYSO, OSA, Florida optical forum, to name a few. His research interests are futuristic tech. in teleportation, photonics, adonics, Mobionics, UPP, Energy, space system, …etc. He received the Honorary Doctorate in Power & Energy from Stanford University, USA.
Two Element Arrays of Circular Patch Antennas in Indoor Clustered MIMO Channels

C. Poongodi¹ P. Arun Saranya² K.V. Archana² A. Shanmugam³

¹Senior Lecturer, ²UG Student, ³Principal, Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam -638401. Email : poongi_cj@yahoo.co.in, saranya_smiling@rediffmail.com, malligaarchana@yahoo.co.in, dras@yahoo.co.in.

ABSTRACT
In this paper we analyze a multiple input multiple output (MIMO) array consisting of two circular microstrip antennas, designed to exploit pattern diversity. The two circular microstrip elements are collocated and stacked upon each other. The spatial correlation coefficients of this array as a function of the mode excited, for realistic clustered MIMO channel models are derived. The angle of arrival/angle of departure are distributed only over the azimuth direction is the assumption taken for circular patch array (CPA) and the performance is compared against an array of two spaced dipoles with the help of the simulation results. Simulation tool used is MATLAB 7.4.

Keywords : Pattern Diversity, Circular Patch Antenna(CPA), Uniform Linear Array(ULA)

1.INTRODUCTION
MIMO technology has attracted attention in wireless communications, since it offers significant increases in data throughput and link range without additional bandwidth or transmit power. MIMO wireless systems use multiple antenna elements at transmit and receive to offer improved capacity over single antenna topologies in multipath channels [1]. In such systems, the antenna channel characteristics play a key role in determining communication performance. The capacity and reliability of MIMO wireless communication systems may be sensitive to the design of the antenna arrays employed at the transmitter and receiver, as well as to the nature of the propagation between the arrays. It has been shown that the capacity of a MIMO system increases linearly with the number of antennas in the presence of a scattering-rich environment [2]. This will ensure that the signals at the antennas in the array are sufficiently uncorrelated with each other [3]. This is where antenna design comes in for MIMO systems.

The throughput that a multiple-input multiple-output (MIMO) channel can support depends on element spacing [4]-[6], array geometry [4],[7],[8], radiation pattern, cross-polarization and the spatial characteristics of the propagation environment (i.e., angle spread, angle of arrival, power angle profile). MIMO antenna arrays can be designed to reduce the channel spatial correlation, resulting in enhanced link performance.

One way to reduce the spatial correlation is to space the antennas far apart by exploiting space diversity. Depending on the spatial characteristics of the MIMO channel, the distance between the array elements needs to be multiple of the wavelength to ensure good system performance. In typical MIMO systems, size and cost constraints often prevent the antennas from being placed far apart. Therefore, space diversity techniques may be insufficient for next generation wireless handsets. One promising solution to overcome the size limitations of wireless devices is pattern diversity. To exploit pattern diversity, the antennas are designed to radiate with
orthogonal radiation patterns as a means to create uncorrelated channels across different array elements [9]. We analyze a MIMO array consisting of collocated circular microstrip antennas. This MIMO array exploits pattern diversity without requiring excessive real estate for spacing the antennas. Different modes can be excited inside the microstrip, yielding different capacity/error rate performance. We consider two element arrays, where the antennas have the same polarization, to isolate the effect of pattern from polarization diversity [10]-[13].

This paper is organized as follows. The general system model and brief overview of MIMO clustered channel model is described in section 2. Spatial correlation coefficients of the CPA are described in section 3. Simulation results in clustered channel model are described in section 4. Finally conclusions are presented in section 5.

2. System Model

In this section, we describe the system model and present a brief overview of MIMO clustered channel model for indoor environments. We model the receive signal of a narrowband MIMO system, with $N_t$ transmit antennas and $N_r$ receive antennas is given by,

$$ y = \sqrt{\frac{\text{SNR}}{N_t}} H x + n $$

$y =$ Received signal.

$\text{SNR}=$ Signal to Noise Ratio.

$N_t =$ Number of transmit antennas.

$x =$ Transmit signal.

$n =$ Zero Mean Additive Gaussian Noise.

$H =$ MIMO channel matrix.

For spatially correlated MIMO channels, the matrix $H$ is generated as [14]

$$ H = \frac{K}{K+1} H_{\text{los}} + \sqrt{\frac{1}{K+1}} H_{\text{nlos}} $$

$$ K =$ Ricean K – Factor.

$H_{\text{los}} =$ Line of sight component of $H$.

$H_{\text{nlos}} =$ Non Line of sight component of $H$.

The LOS component of the channel is assumed to be rank one and from [15],[16] it is given by,

$$ H_{\text{los}} = a(\Omega_{\text{los},t}) \cdot a(\Omega_{\text{los},r})^H $$

$a(\Omega) =$ Array response as a function of solid angle $\Omega = (\varphi, \theta)$.

$\Omega_{\text{los},t}$ and $\Omega_{\text{los},r}$ are the Angle of Departure (AoD) / Angle of Arrival (AoA) corresponding to the LOS component at the transmitter and receiver sides, respectively.

The NLOS channel matrix from [4],[17] is defined as ,

$$ H_{\text{nlos}} = R_t^{1/2} H_n R_r^{1/2} $$

$R_t =$ Transmit Spatial correlation matrix.

$R_r =$ Receive Spatial correlation matrix.

$H_n =$ Matrix of complex Gaussian fading coefficients.

$N =$ Number of array elements.

This Geometry of the clustered channel model representing clusters and propagation paths is as shown in the fig.1.

![Figure 1: Geometry Of The Clustered Channel Model Representing Clusters And Propagation Path](image-url)
The angle $\phi_c$ is the mean AoA of the cluster, and $\phi$ is the AoA offset of the propagation path [9]. In clustered channel models, the scattering objects around the transmit/receive arrays are modeled as clusters. The entries of the spatial correlation matrix are a function of the transmit/receive array and the spatial characteristics of the MIMO channel. Here it has considered only the azimuth directions.

3. CIRCULAR PATCH ARRAY AND UNIFORM LINEAR ARRAY

A. Circular Patch Array

The properties of circular microstrip antennas have been studied in [18]-[20]. In [18], it was shown that by exciting different modes of circular patch antennas, it is possible to obtain different radiation properties. In addition, by varying the size of the antennas as well as the feed location, different polarizations and radiation patterns can be generated in far-field. The orthogonality of the radiation patterns of circular patch antennas as a means to reduce correlation between the diversity branches of the MIMO array is used here.

The electric field of the $n^{th}$ mode excited inside the circular patch antenna as a function of its $\theta$ and $\phi$ far-field components from [9] are expressed as,

$$E^{(n)}(\phi,\theta) = -e^{j\pi/2} \frac{V_0^{(n)}}{2} k_0\rho (J_{n+1}(k_0\rho) + J_{n-1}(k_0\rho)) \cos\theta \sin(\phi - \theta_0).$$

where,

$$V_0^{(n)} = \text{Input Voltage.}$$

$$k_0 = \text{wave number.}$$

$$J_n = J_n(k_0\rho \sin \theta) = \text{Bessel functions of the second kind.}$$

$$n = \text{order.}$$

$\rho = \text{Radius of the microstrip antenna.}$

$\theta_0 = \text{Reference angle corresponding to the feed point of the antenna.}$

To isolate the effect of pattern from space diversity, the patch antennas are assumed to be collocated and stacked on top of each other [19]. The same mode is excited for both elements of the MIMO array, and tunes the phase to produce orthogonal radiation patterns across the diversity branches. For the case of a two-element MIMO array, we feed one antenna with $\phi_0^{(1)} = 0$, and the other antenna with $\phi_0^{(2)} = \pi / (2n)$. As result of this feeding technique, we get orthogonal radiation patterns for any mode excited within the antennas. Mode 0 is discarded since it does not yield any pattern diversity, due to its isotropic radiation pattern over the azimuth directions.

The array response of the CPA from [4] is given by,

$$a_{cpa} (\phi) = \gamma(\rho, n) [\cos(n \phi), \sin(n \phi)]^T$$

where,

$$\gamma(\rho, n) = e^{j\pi n/2} \frac{V_0^{(n)}}{2} k_0\rho (J_{n+1}(k_0\rho) + J_{n-1}(k_0\rho))$$

$\phi = \text{azimuth AoA / AoD.}$

The overall power radiated by the array is fixed to be a constant for any mode. Assume,

$$\sum_{n=0}^{\infty} |a_{cpa}(\phi)|^2 = 1$$

for any azimuth direction.

B. Uniform Linear Array

The performance of the CPA against a ULA is compared, and measured the gains achievable through a pattern over space diversity techniques. As for the CPA, it has been considered the case of a ULA with two elements, consisting of half-wavelength dipole antennas vertically polarized, with variable element spacing. It has been expressed the array response of this ULA as,

$$a_{ula} (\phi) = [1, e^{j\rho_0 \sin \phi}]^T$$
The power radiated by any array is a combination of the power radiated by each element, which depends on the input power through the reflection coefficients [21]. The conditions of constant "radiate" power is equivalent to constant “input” power between two arrays, as long as the reflection coefficients for the two arrays are the same.

C. Spatial Correlation Coefficients of the CPA

The spatial correlation coefficients of the CPA, assuming the MIMO clustered channel model is computed. The voltage received at the port of the lth patch from [22] is given by,

$$\int \pi 4 (\theta, \phi) E(l) d\Omega$$

This correlation coefficient is derived for the single-cluster channel, but its expression can be easily extended to multiple clusters by adding up the correlation coefficients due to each of the clusters, because of the independency of the clusters.

The cross correlation coefficients of the CPA is given by,

$$\gamma_{12}(\theta_1, \theta_2, \phi_1, \phi_2) = \frac{|\gamma(\rho_1, \rho_2)|^2 \sin(2\phi_1 \phi_2)}{1 + 2(n \sigma_\phi^2)^2}$$

D. Eigenvalues Of The Spatial Correlation Matrix

The eigenvalues of the correlation matrix for the CPA is given by,

$$\lambda_{\hat{e}} = 1 \pm \frac{1}{1 + \frac{(kd \cos \phi)^2}{\sigma_\phi^2}}$$

As the mode number increases, the eigenvalues become closer to one, and the product \( \lambda_1 \lambda_2 \) is maximized. For high n, the radiation pattern of the circular patch antenna is characterized by a large number of lobes, which yields high pattern diversity. The approximate expression of the correlation coefficients is used for ULAs with a Laplacian distributed power azimuth spectrum and is given by,

$$\gamma_{12} = 1 \pm \frac{1}{\sigma_\phi^2 (kd \cos \phi)^2}$$

E. Channel Capacity

The tight upper bound to the ergodic capacity for spatial multiplexing (SM) systems (with equal power allocation across the transmit antennas) is considered. We assume zero mean single-sided (only at the transmitter) correlated MIMO channels. This upper bound is expressed as [9]
Two Element Arrays of Circular Patch Antennas in Indoor Clustered MIMO Channels

\[ C \leq \log_2 \left[ 1 + 2SNR + \frac{SNR^2}{2} \lambda_1 \lambda_2 \right] \]  

(15)

\( \lambda_1 \) and \( \lambda_2 \) are the eigenvalues of the spatial correlation matrix \( R \). The capacity increases as a function of the product of the eigenvalues \( (\lambda_1 \lambda_2) \).

4. RESULTS AND DISCUSSION

Auto Correlation Coefficient for the first and second antenna of the CPA is shown in fig.2 and fig.3 respectively. Fig.4 shows the cross correlation Coefficient of the two antennas of the CPA (\( r_{12} \)). If the mode number (\( n \)) increases, the frequency of the oscillations of the auto correlation and the cross correlation increases. This is due to the higher number of lobes in the radiation pattern for the higher order modes. The amplitude of the oscillations decreases for increasing mode number.

Fig.5 shows Eigenvalues of the correlation matrix for CPA: mode=3. When the mode number increases, the eigen values become closer to one and the product of \( \lambda_1 \) and \( \lambda_2 \) is maximized. This is due to the higher decorrelation between the diversity branches of the CPA for higher order modes. This leads to the improved channel capacity and error rate performance.

For high values of mode number, the radiation pattern of the CPA is characterized by a large number of lobes which yields high pattern diversity. Capacity increases as the function of the product of the eigen values (\( \lambda_1 \) \( \lambda_2 \)). For mode 1, the CPA outperforms the ULA only for angles close to the end fire directions. For higher order modes, the CPA always provides better performance than the ULA. Fig.6 shows ergodic capacity for the CPA and
ULA (d = 0.5 λ). The maximum capacity of the CPA is close to its saturation point when mode 3 is employed. Higher the mode number, larger the size of the microstrip antenna for the fixed dielectric constant of the substrate and carrier frequency.

5. CONCLUSION

In this paper we analyzed the performance of the MIMO system with the help of Circular Patch Array for different modes. The correlation is getting increased with lower order modes. Hence it is not suitable for higher capacity requirements. In the simulation results, the circular patch array with two elements is analyzed which exploits pattern diversity. The frequency of the oscillation for the correlations is also getting increased with increase in mode number. For reducing the correlation and in order to increase the capacity the higher order modes are suitable.

REFERENCES


Two Element Arrays of Circular Patch Antennas in Indoor Clustered MIMO Channels


Implementation of FPGA Based Incremental PID Controller Using Conventional Method and Distributed Arithmetic Algorithm

Mariamma John¹, B. Thilagavathi², Anitha Mary X³

ABSTRACT

In this paper, two efficient design schemes for implementation of the Incremental Proportional-Integral-Derivative (PID) controller using Field Programmable Gate Array (FPGA) technology are presented. Conventional implementation and also Distributed arithmetic based implementation of Incremental PID controller is done. Conventional implementation contains a large number of multipliers and adders. It’s not focused on optimal use of hardware resources and do not efficiently utilize the memory-rich characteristics of FPGAs. An FPGA chip consists of a lot of memory blocks, referred to as Look-Up Tables (LUT), which can be utilized to implement efficient designs. In the Distributed Arithmetic (DA) scheme is an efficient LUT design method, and is very promising in the FPGA implementation of PID controller. Distributed arithmetic replaces multiplication by addition and shifting. The values are precomputed and placed in LUT. Based on the LUT scheme, the proposed PID controller reduces the cost of the FPGA design by enabling the chip to accommodate more logic and arithmetic functions while requiring less power consumption.

Keywords : PID Controller, Incremental Form, Conventional Method, Distributed Arithmetic, FPGA.

1. INTRODUCTION

Proportional Integral Derivative (PID) controller is the most common type of controller used in dynamic systems. An important feature of this controller is that it does not need a precise analytical model of the system that is being controlled. For this reason, PID controllers have been widely used in process control, manufacturing, robotics, automation, transportation, and interestingly in real-time scheduling of concurrent tasks in multi-tasking applications.

PID controllers are often combined with logic, sequential functions, and other blocks to implement complicated systems. Implementation of PID controllers has gone through several stages of evolution, from the early mechanical and pneumatic designs to the microprocessor-based systems. Recently, Field Programmable Gate Arrays (FPGA) have become an alternative solution for the realization of digital control systems, previously dominated by the general purpose microprocessor and application specific integrated circuits (ASIC). The FPGA-based controllers offer advantages such as high-speed computation, complex functionality, real – time processing capabilities, and low power consumption. These are attractive features from the embedded systems design point of view.
Conventional implementation of FPGA based controllers have not focused on optimal use of hardware resources. These designs usually require a large number of multipliers and adders and do not efficiently utilize the memory-rich characteristics of FPGAs. An FPGA chip consists of a lot of memory blocks, referred to as Look-Up Tables (LUT), which can be utilized to implement efficient designs. In this work, we utilize the Distributed Arithmetic (DA) scheme, which is an efficient LUT design method, and is very promising in the FPGA implementation of PID controller. Incremental form of PID controller is obtained from the basic controller equation.

The organization of this paper is as follows. In section 2, PID controller is discussed. In Section 3, incremental form of PID is discussed. Section 4 throw some light on multiplier based conventional method. Section 5 and 6 tells about DA and its fusion on incremental equation. Simulation results are discussed in section 7.

2. PID CONTROLLER

The PID control algorithm is one of the most commonly used control algorithms in industry. A proportional–integral–derivative controller (PID controller) is a generic control loop feedback mechanism (controller) widely used in industrial control systems. A PID controller attempts to correct the error between a measured process variable and a desired set point by calculating and then outputting a corrective action that can adjust the process accordingly and rapidly, to keep the error minimal as shown in Fig 1.

In Fig 1 variable (e(t)) represents the tracking error, the difference between the desired input value (set point) and the actual output (output). This error signal (e(t)) will be sent to the PID controller, and the controller computes both the derivative and the integral of this error signal. The signal (u(t)) just past the controller is now equal to the proportional gain (k_P) times the magnitude of the error plus the integral gain (k_I) times the integral of the error plus the derivative gain (k_D) times the derivative of the error. T_I is the reset time and T_D is the derivative time.

\[
    u(t) = k_p e(t) + \frac{1}{T_i} \int e(t) dt + T_d \frac{de(t)}{dt}
\]

The PID controller calculation (algorithm) involves three separate parameters; the proportional, the integral and derivative values. The proportional value determines the reaction to the current error, the integral value determines the reaction based on the sum of recent errors, and the derivative value determines the reaction based on the rate at which the error has been changing. The weighted sum of these three actions is used to adjust the process via a control element such as the position of a control valve or the power supply of a heating element.

3. INCREMENTAL FORM

Analog Control refers to the design and implementation of controllers in the continuous domain. There are a number of ways by which this common and versatile controller can be implemented in discretised form. The two forms used are incremental (velocity) form and positional (full valve) form. The equation (1) is discretised and the following equation is obtained:

\[
    u[n] = k_p \Delta y[n] + k_i \sum_{j=0}^{n} \Delta y[j] + k_d [\Delta y[n] - \Delta y[n-1]]
\]

Where

- k_I is the integral coefficient and
- k_D is the derivative coefficient. This form is called

![Figure 1: Block Diagram of a Control System](image-url)
the position form of the PID algorithm. An alternative would be to compute $u[n]$ based on past output $u[n-1]$ and correction term $\Delta u(n)$. This approach is often called as the velocity form of the PID algorithm. The first step in this regard would be to calculate $u[n-1]$ based on equation (2).

$$u[n-1] = k_p e[n-1] + k_i \sum_{j=1}^{n-1} e[j] + k_d \left[ e[n-1] - e[n-2] \right]$$ (3)

Then calculate correction term as:


Where

$$k_e = k_p + k_i + k_d$$
$$k_f = -k_p - 2k_i$$
$$k_d = k_d$$

The current control output is calculated as:

$$u[n] = u[n-1] + \Delta u[n]$$ (5)

$$= u[n-1] + k_e e[n] + k_f e[n-1] + k_d e[n-2]$$ (6)

The above equation (6) represents the incremental form of PID controller.

4. **Conventional Method**

Incremental form of PID is decomposed into its basic operations. Here $p$ and $p_i$ refers to the controlled variable and its desired value (setpoint) respectively. $p_0,p_2,p_2,s_1,s_2$ are temporary variables.

$$e[n] = P + (-Pd)$$

$$P0 = k_p e[n]$$

$$P1 = k_i e[n-1]$$

$$P2 = k_d e[n-2]$$

$$S1 = P0 + P1$$

$$S2 = P2 + u[n-1]$$

$$u[n] = S1 + S2$$

Input represents the current position $P$. The negation of $P$, $P_{neg}$, is generated by bit-wise complementing and adding 1. At the rising edge of control, signal $e[n]$ of the last cycle is latched at register REG1, thus becomes $e[n-1]$ of this cycle. In the same manner, $e[n-2]$ and $u[n-1]$ are recorded at REG3 and REG4 by latching $e[n-1]$ and $u[n]$ respectively. The registers can be set to initial values of 0 by asserting the reset signal, reset. As long as the desired position $P_d$ is also initialized to 0 when the system is reset, the control output is 0, which can keep the controlled device (i.e. the motor, in this system) static. The computed control output $u[n]$ may exceed the range that the controlled device can bear. Bounder is a value limitation logic that keeps the output in the user defined range of $UpBound$ and $LowBound$.

The VHDL coding for conventional method were written and simulation result was obtained.

5. **Distributed Arithmetic**

Distributed Arithmetic (DA), along with Modulo Arithmetic, are computation algorithms that perform multiplication with look-up table based schemes. Both stirred some interest over two decades ago but have languished ever since. Indeed, DA specifically targets the sum of products (sometimes referred to as the vector dot product) computation that covers many of the important DSP filtering and frequency transforming functions. Ironically, many DSP designers have never heard of this algorithm. Inspired by the potential of the Xilinx FPGA look-up table architecture, the DA algorithm was resurrected in the early 90’s and shown to produce very efficient filter designs.

The derivation of the DA algorithm is extremely simple but its applications are extremely broad. The mathematics includes a mix of Boolean and ordinary algebra and requires no prior preparation - even for the logic designer.
The arithmetic sum of products that defines the response of linear, time-invariant networks can be expressed as:

\[ y(n) = \sum_{k=1}^{K} A_k x_k(n) \]  

(7)

where:

- \( y(n) \) = response of network at time \( n \).
- \( x_k(n) \) = \( k \)th input variable at time \( n \).
- \( A_k \) = weighting factor of \( k \)th input variable that is constant for all \( n \), and so it remains time-invariant.

In filtering applications the constants, \( A_k \), are the filter coefficients and the variables, \( x_k \), are the prior samples of a single data source (for example, an analog to digital converter). In frequency transforming whether the discrete Fourier or the fast Fourier transform - the constants are the sine/cosine basis functions and the variables are a block of samples from a single data source. Examples of multiple data sources maybe found in image processing.

The multiply-intensive nature of equ 7 can be appreciated by observing that a single output response requires the accumulation of \( K \) product terms. In DA the task of summing product terms is replaced by table look-up procedures that are easily implemented in the Xilinx configurable logic block (CLB) look-up table architecture.

We start by defining the number format of the variable to be 2’s complement, fractional - a standard practice for fixed-point microprocessors in order to bound number growth under multiplication. The constant factors, \( A_k \), need not be so restricted, nor are they required to match the data word length, as is the case for the microprocessor. The constants may have a mixed integer and fractional format; they may be written in the fractional format as shown in equ. 8

\[ x_k = x_{k0} + \sum_{n=1}^{n_{max}} x_{kn} 2^{-n} \]

where \( x_{k0} \) is a binary variable and can assume only values of 0 and 1. A sign bit of value -1 is indicated by \( x_{k0} \). Note that the time index, \( n \), has been dropped since it is not needed to continue the derivation. The final result is obtained by first substituting equ.8 into equ.7

\[ y = \sum_{n=1}^{K} A_k \left[ x_{k0} + \sum_{n=1}^{n_{max}} x_{kn} 2^{-n} \right] = \sum_{n=1}^{K} A_k x_{k0} + \sum_{n=1}^{K} \sum_{n=1}^{n_{max}} A_k x_{kn} 2^{-n} \]

and then explicitly expressing all the product terms under the summation symbols equation 9:

Each term within the brackets denotes a binary AND operation involving a bit of the input variable and all the bits of the constant. The plus signs denote arithmetic sum operations. The exponential factors denote the scaled contributions of the bracketed pairs to the total sum. You can now construct a look-up table that can be addressed by the same scaled bit of all the input variables and can access the sum of the terms within each pair of brackets.

The arithmetic operations have now been reduced to addition, subtraction, and binary scaling. With scaling by negative powers of 2, the actual implementation entails the shifting of binary coded data words toward the least significant bit and the use of sign extension bits to maintain the sign at its normal bit position. The hardware implementation of a binary full adder (as is done in the CLBs) entails two operands, the addend and the augend to produce sum and carry output bits. The ultimate in
gate efficiency would be a single DALUT, a single parallel adder, and, of course, fewer flip-flops for the input data source. Again with our B=16 examples, a rephrasing of equation 9 yields the desired result:

\[ y = \{ \left( \dagger \sum_{15} \right) \left( \dagger \sum_{14} \right) \ldots \left( \dagger \sum_{0} \right) \} \times s \]

Starting from the least significant end, i.e. addressing the DALUT with the least significant bit of all K-1 and then added to the DALUT input variables the DALUT contents, \([\sum_{15}]\), are stored, scaled by 2^{-1} contents, \([\sum_{14}]\) when the address changes to the next-to-the-least-significant bits. The process repeats until the most significant bit addresses the DALUT, \([\sum_{0}]\). If this is a sign bit a subtraction occurs. Now a vision of the hardware emerges. A serial shift register, B bits long, for each of the K variables addresses the DALUT least significant bit first. At each shift the output is applied to a parallel adder whose output is stored in an accumulator register. The accumulator output scaled by 2^{-1} henceforth, the adder, register and scalar shall be referred to as a scaling accumulator. The functional blocks are shown in fig. 2. All can be readily mapped into the Xilinx 4000 CLBs. There is a performance price to be paid for this gate efficiency - the computation takes at least B clocks.

6. DA ON INCREMENTAL FORM

Fig 3 shows PID based control system; here the output from ADC will be 8 bit if we are using an 8 bit ADC. This should be multiplied with step size and then only we get P. The difference between P and P_d gives error e[n] which forms the input to controller and it produces a desired controller output u[n].

![PID Based Control System](image)

We need to create 2 LUT for implementing PID with distributed arithmetic. A LUT can be of 2^k x 1 size, where k is the number of inputs to LUT.

LUT, for the ADC Side

Here we have to find P by the following equation

\[ P = a[n] \times s \]  

Where, \( a[n] \) is the n bit output from ADC \( s \) is the step size

so for the LUT we have 1 input \( a[n] \), so we form 2^1 x 1 LUT. The contents of LUT1 table are as shown in Table 1.

<table>
<thead>
<tr>
<th>( a[n] ) input</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>s</td>
</tr>
</tbody>
</table>

Fig 4 shows the block diagram of ADC side to implement. The \( a[n] \) output from ADC acts as the input to LUT1. It will be placed in a parallel serial register and will be giving out 1 lsb at 1 clock this become the input to LUT. In LUT value corresponding to the input bits ‘0’ and ‘1’
Implementation of FPGA Based Incremental PID Controller Using Conventional Method and Distributed Arithmetic Algorithm

is pre-computed and stored so as the input bit come corresponding output will be given out from LUT which will be given to adder and then to accumulator and shifted by $2^{-1}$. After shifting its fed back to adder/subtractor and gets added with next output of LUT. Thus after $n$ clock cycles the result of multiplication (actually done by shifting and addition) is obtained in $P$.

Figure 4 : Block Diagram of ADC Side

$LUT_2$ for error containing part of incremental form

Equation (6) can be written as shown below

$$u[n] = u[n-1] + E$$  \hspace{1cm} (4.14)

Where, $u[n]$ is the controller output

$u[n-1]$ is the past controller output

$$E = k_p e[n] + k_i e[n-1] + k_d e[n-2]$$  \hspace{1cm} (4.15)

Here $e[n]$, $e[n-1]$, $e[n-2]$ are the inputs to LUT. Size of LUT will be $2^3 	imes 1$. The contents of $LUT_2$ table are as shown in Table 2. As in Fig 5 the input $P$ (which comes as the output of $LUT_1$ after $n$ clock cycles) is subtracted from $P$, and it form the error $e[n]$ which will put in parallel serial register and will output 1 lsb at 1 clock.

Figure 5 : Block Diagram of Incremental Equation Side

At the rising edge of control, signal $e(n)$ of the last cycle is latched at register $reg1$, thus becomes $e(n-1)$ of this cycle. In the same manner, $e(n-2)$ and $u(n-1)$ are recorded at $reg2$ and $reg3$ by latching $e(n-1)$ and $u(n)$ respectively.

At first instant $e[n-1]$ and $e[n-2]$ are zero. $e[n], e[n-1], e[n-2]$ forms the input of $LUT_2$ and as said before the output gets added and shifted by $2^{-1}$. And after $n$ clock cycles we get the $u[n]$ which is added with previous $u[n-1]$ to give the output of controller.

7. SIMULATION RESULTS

In the conventional method’s coding output from 8 bit ADC is multiplied with the step size. This output of this multiplication forms $P$ input which will be stored as ‘$Pi’ containing fractional part and ‘$Pd’ containing integer part.

Its twos compliment is taken an added with the desired output ‘$Pd’ the fractional part of ‘$P’ and ‘$Pd’ is first added and if it produces a carry it will be added with their integer portions. When reset zero initial values of $e[n]$, $e[n-1]$, $e[n-2]$, $u[n-1]$ are made zero. The output value will be obtained only when clock is given.

The error obtained is stored as separate integer and fraction portion. The value of $e[n]$ will be latched to $e[n-1]$ and $e[n-2]$ as explained previously. Calculations are done taking fractional and integer portion separately.

The fraction part of output is obtained in ‘outputf’ and the integer portion in ‘outputi’.
The RTL schematic of code is as shown in Fig 6, the multipliers and adders are clearly seen on the diagram. The simulation wave form is shown in Fig.7. Conventional implementation of FPGA based controllers have not focused on optimal use of hardware resources. These designs usually require a large number of multipliers and adders and do not efficiently utilize the memory-rich characteristics of FPGAs. So we now go for Distributed Arithmetic scheme which uses LUTs of FPGA.

8. CONCLUSION

In this paper conventional method of incremental pid controller has been simulated in model sim, now going to fuse distributed arithmetic algorithm onto this incremental equation. By proposed DA-based LUT scheme, the memory inside FPGA has been utilized to provide efficient design for PID controllers.

Future works include writing code for a DA scheme based on the block diagram designed. Implementation of algorithms on FPGA SPARTAN 2E and hardware interface like a temperature control system.

REFERENCES


[5] “The role of distributed arithmetic in FPGA based signal processing”
Author's Biography

Mr. K. Rajsekaran, received B.E. ECE in PSG tech, Coimbatore in the year 1976 and M.E (Instrumentation) from IIT Kharagpur in 1978. Currently pursuing Ph.D in Anna University Coimbatore. He is working as HOD for the past 7 years.

Mariamma John, P.G. scholar from Karunya university, Coimbatore.

Anitha Mary. X, received Bachelor of Engineering (electronics and instrumentation engineering) from Karunya Institute of Technology in the year 2001 and Master of Engineering (VLSI Design) from Anna University, Coimbatore in the year 2009. Currently pursuing Ph.D in Control System area. She attended 3 National and 4 International Conference. She guided many U.G. and P.G. projects. Currently working as Assistant Professor in the EIE department.