CHAPTER 1

INTELLIGENT INTEGRATED POSTAL AUTOMATION: AN INTRODUCTION

1.0 PREAMBLE

The world is witnessing rapid changes in lifestyle with many countries embracing globalization and liberalization. Every system and process in almost all walks of life is being overhauled to make it more efficient. The prime mover for such a change in various systems and processes is the advancement in Information Technology leading to automation of systems and processes. Automation is the art of making processes or machines self acting or self-moving [W-1]. Automation is a step beyond mechanization, while mechanization provided human operators with machinery to assist them to meet the physical requirements of work; automation greatly reduces the need for human intervention in completion of the work [W-2]. The purpose of automation of any process or system is to make it more efficient and flexible.

The automation of a process or system is achieved by devising microprocessor-controlled machines such as computers, programmable logic controllers, embedded systems etc, for performing the required tasks of the process. The automation is implemented by programming the microprocessor to carry out the jobs sequentially or on occurrence of certain events. The availability of highly sensitive sensors and transducers has aided the automation of complex industrial processes. The processes requiring evaluation of complex mathematical systems can also be easily automated using appropriate software tools and the large computing power available with processors.

The automation of processes/ systems has been successful when automation is used to substitute the mechanical/ repetitive works of human beings. Moving objects from one location to another using robots, starting or stopping an activity when a predefined condition is met for example opening a solenoid valve to start a conveyor belt when an...
object is deposited on it, performing large computations for applications as diverse as weather forecasting, nuclear fusion simulation, psephology, etc, are some examples where automation has resulted in efficient processes.

In comparison, processes or systems, which involve human expertise and intelligence, are difficult to automate. More specifically it is very difficult for the machines to emulate the object recognition, language recognition and interpretation ability of human beings. New developments in machine intelligence, pattern recognition and soft computing technologies such as genetic algorithms, fuzzy systems, neural networks, symbolic data analysis have provided tools through which emulation of intelligent human tasks can be attempted. Use of these technologies has resulted in development of better automation systems such as fuzzy controlled washing machines, microwave ovens etc. These tools can also be employed in automation of social, financial and economic systems and processes. Banking systems, airline reservation and railway reservation are some systems, which are greatly benefited from automation using the modern tools.

Automation is a bland process that removes the human element as far as possible. For the automated system to be efficient the manual process/ system, which is automated, has to be efficient. Bill Gates says The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency [W-3]. The postal system that offers written mail conveyance from sender to the receiver is one such efficient system, which has evolved over the years. The rationale of this research is to develop computational strategies aimed to devise an Intelligent Integrated Automation System for Efficient Processing of Postal Mail.

Communication using written mail transfer is the oldest form of formal communication known to mankind. This type of communication continues to thrive in the modern day and is a necessary service for transfer of written information and original documents till date. Section 1.1 describes the postal system and brings out the need for its automation
1.1 INTRODUCTION TO POSTAL SYSTEM AND THE NEED FOR AUTOMATION

The word 'post' stands for official conveyance of letters, parcels, and other items, called mail, as per the Oxford Dictionary. Every country has setup an organization to handle the conveyance/ transport of mail from one place to another. In most of the countries this organization is either a Government department or an autonomous body. Many of these organizations have come together to set up a postal union called Universal Postal Union (UPU). The Universal Postal Union is the second oldest international organization, established in 1874 and headquartered in the Swiss capital of Bern [W-4]. Mail is exchanged across international boundaries through the universal postal union.

Mail conveyance from the sender to the receiver is a very labor intensive process made up of many sub tasks. The mail deposited in the letterbox by the sender is collected, sorted according to their destinations and sent towards the destination in a mailbag. In the transit, the mail might reach other intermediate places, where sorting is again carried out. Once the mail reaches it’s destination it is delivered to the addressee after ascertaining the destination address written on the mail. The complete mail conveyance process is depicted in Figure 1.1. The type of mail conveyance described above is sometimes called Snail Mail, due to the time taken for delivery of the mail to the recipient or Surface Mail as the mail is transported physically and delivered to the recipient. With the onset of globalization and privatization, many private agencies have also taken up the job of mail delivery to the society and are providing tough competition to the official/ designated postal service providers, although in some countries the official postal service providers have monopoly over conveyance of some type of mail.

Along with the competition from private agencies that provide mail conveyance service, the postal service faces threat from various forms of electronic communication enabled by advances in computer and communication technology. In the Internet centred world of today the modern communication techniques like email, fax, telex, etc, have drastically reduced the time required for communication (making communication instantaneous). The advances in electronic communication technology have brought down the cost of
communication. In comparison the cost of written mail conveyance is increasing because transportation costs are increasing and skilled/unskilled human labor required to complete the various sub tasks of mail conveyance is at a premium. In spite of all this the need for snail mail, or postal mail cannot be totally eliminated and is still a thriving and sought after medium of communication. This is because some people like writing letters and there is still a need to deliver documents in their original form that need mail conveyance. Furthermore many companies send catalogues, journals, and magazines by post. Kubota and Egami (1999) observe that with the changing social environment, the type of mail that is delivered has also undergone a considerable change. Thirty years ago more than half of
Intelligent Integrated Postal Automation: An Introduction

the mail was private letters from individual to individual. With the development of global economy and the change in the social system, mail from enterprises, who are large volume senders to an individual has drastically increased. Now about 70-80% of the mail is from enterprises to an individual and this has given rise to special postal service referred to as Direct Mail Service [W-4]. According to the statistics available with UPU (Universal Postal Union) more than 435 billion mail items are processed and delivered each year to all corners of the world. Hence there will be continued use of the postal system in the foreseeable future.

Also, some of the constraints of technology such as requirement of the knowledge of English, the exorbitant cost of implementing the last mile of electronic communication in an unfriendly terrain and remote places, has limited the use of modern communication tools such as email, fax etc, to a privileged few, who are on the right side of digital divide. In such a scenario the written postal mail conveyance is the only affordable and feasible form of communication. The features of postal system that cannot be matched by other forms of communication are,

- Security
- Language independence
- Reach to the remotest/ farthest place
- The presence of a friendly postman etc.

The Indian Postal System founded in 1764 will be taken as case study in this research work. The Indian Postal System is the most widely distributed post office system in the world with more than 155,000 post offices through out India [W-5]. The postal system being the only feasible communication media for certain under privileged people is more true in the Indian context, where the digital divide is manifest. A brief description of Indian Postal System is given in appendix to chapter 1.

As the mail conveyance / postal service is a necessary service there is a need to make it more efficient so as to meet the expectation of its users and face the competition from the various electronic communication tools. The postal mail service provided either by the official designated organisation or private agencies can be made efficient by automating
Intelligent Integrated Postal Automation: An Introduction

the subtasks of postal service/ mail conveyance. Postal automation aims at developing computational strategies for transfer of mail from sender to receiver quickly reliably and economically. Nagabhushan (1998) has enlisted the various tasks of postal service that are to be automated especially in the Indian context. The tasks of mail conveyance that can benefit from automation include mail collection routing, mail delivery sorting and optimal route generation for mail delivery. Section 1.2 explores the various challenges/ issues in automating the postal service.

1.2 ISSUES IN POSTAL AUTOMATION

The automation of various tasks associated with mail conveyance/ processing requires automatic address reading, validation of addresses, identifying the mail delivery point, generation of optimal paths for distribution of mail, collection of mail from letter-boxes and other allied tasks. The tasks associated with mail conveyance can be grouped into three major categories, namely mail collection, mail sorting and mail delivery/ distribution. The sequence of these tasks in mail conveyance is depicted in Figure 1.2. Mail collection involves gathering mail from letter boxes, mail sorting involves classifying and grouping the mail for forwarding it towards the destination or for delivery to the addressee. The mail delivery/ distribution constitutes the delivery of mail to the doorstep of the addressee.

![Figure-1.2: Subtasks of Mail Conveyance](image)

Each of the tasks is made up of many different subtasks, which can be automated to make them efficient. In the following paragraphs we explore the various such subtasks in postal mail processing.

The **mail collection** can be made efficient by fixing optimal and convenient locations for locating the letter boxes using operation research techniques. Also optimal mail collection routes can be found using arc routing techniques (Ahr, 2000).
Mail sorting is the most important task of postal mail conveyance and requires skilled personnel. Automation of the various subtasks of mail sorting is the major challenge in postal automation. The various jobs involved in mail sorting are depicted in Figure 1.3.

The jobs of mail sorting can be further classified into two major categories namely computer vision tasks and knowledge based tasks. The computer vision tasks involve identifying the destination address block, segmenting the address block and recognizing the components after script identification. The computer vision tasks are enlisted in blocks 1,2 & 3 of Figure 1.3. The knowledge-based tasks include address interpretation, address translation if required, address component labeling, address validation, delivery point mapping, etc. All these tasks require emulation of human expertise with the help of accumulated knowledge. The knowledge-based tasks to make the mail sorting efficient are depicted in blocks 4,5,6,7 & 8 of Figure 1.3.

The mail distribution task can be made efficient by devising techniques for automatically dividing the mail distribution task amongst postmen and finding the best and practically feasible traversal path for distribution using optimization and soft computing techniques.

In summary the issues in automation of mail conveyance/processing so as to make postal mail processing efficient are as listed below:
i) Generation of mail collection routes

ii) Identification of the face of the mail on which destination address is written (Facing and Canceling of mails)

iii) Identification of the Location of Destination Address Block (DAB) and its extraction

iv) Division/segmentation of address (Destination Address / Source Address) into lines/components.

v) Identification of class of the component (machine printed/ hand printed/ cursive) along with the script/language (especially in a multilingual/multi script country like India)

vi) Reading of address in the context of local knowledge

vii) Validation of the recognized address based on available local knowledge

viii) Mapping of mail address to mail delivery point (MDP)

ix) Generation of mail distribution graphs on a given day using the information about the mail to be delivered on the day

x) Generation of best practically feasible path for distribution of mail considering various optimization parameters

xi) Information mining from Geographical Information System (GIS) and mail data.

In addition to these there are other related tasks namely stamp/article verification/validation which require the attention of researchers. The issues ii-vi are basically computer vision (document image analysis) problems and many works are reported in literature, whereas issues vii and viii are knowledge based tasks and have not attracted much attention of researchers in postal automation area. Operation research groups have tackled the issues i and ix to xi, as routing problems and very few works are reported on the GIS application for postal information. Literature survey relevant to postal automation is presented in the section 1.3.
1.3 LITERATURE SURVEY RELEVANT TO POSTAL AUTOMATION

The literature survey reveals that researchers around the world are addressing various issues related to building of subsystems required for postal automation. NEC Inc, Japan [W-6], Siemens AG [W-7] Germany, Stanford Research Institute, USA [W-8] are some of the companies that are promoting postal mechanization/automation research in a big way. The machines for pre-processing, facing and canceling with sufficient degree of sophistication are commercially available. Kubota and Egami, (1999) have introduced three generations of postal mechanization and have proposed a carrier (delivery) sequence sorting as the third generation of mechanization as applied to Japanese handwritten mails. Osamu et al, (1999) have proposed a new mail processing system based on computer networks, address reading and bar coding techniques. Masashige et al, (1999) provide a description of mechanization of postal system in various countries such as Taiwan and Singapore. Tadashi et al, (1999) describe a video coding system employed in the postal automation system developed by NEC. Yasuharu et al, (1999) propose an OCR to read addresses printed in Finnish language. Katsuhiro et al, (1999) describe the formats of postal destination and ID-Tag barcodes used in Japan. Invisible bar code and bar code recognition methodology is introduced in (Tesuo et al, 1999). Yoshikazu et al, (1999) introduce a Management Information System (MIS) for postal mail processing. Hiroshi and Yusushi (1999) introduce the two-dimension barcode and stamp recognition methodology.

Various issues in postal automation are being addressed in the Center for Excellence in Document Analysis and Recognition (CEDAR) at Buffalo University, USA [W-9]. The Department of Studies in Computer Science, University of Mysore has also taken up Postal Automation work in the Indian Context as one of its thrust areas of research (Nagabhushan, 1998; Premalatha and Nagabhushan, 2001; Nagamani and Nagabhushan, 2003). Nagabhushan (1998) has described the various tasks in Indian Postal Services that will benefit from automation including mail collection routing to mail sorting and optimal route generation for mail delivery. Premalatha and Nagabhushan (2001) have suggested an algorithmic prototype for verification and validation of PINCODE, the paper also includes
algorithms for address block location, numeral reading and an expert system based verification of PINCODE. Nagamani and Nagabhushan (2003) have proposed a lexicon-based approach for destination postal code generation using the postal sorting and geographic knowledge. Roy et al, (2005) propose a system for Indian Postal Automation, the paper presents algorithms for destination address block (DAB) location, script identification, numeral recognition for PINCODE reading and word recognition for city name reading taking Bangla and English scripts into account.

Researchers around the world have proposed different approaches to various issues related to postal automation. The works related to document image analysis and text reading, knowledge based systems and optimization of arc routes have direct relevance to postal automation. In the following paragraphs a few prominent works in these areas are briefly summarized. The detailed literature survey is provided in later chapters.


Jain and Yu (1998) have proposed a bottom up approach based on connected component extraction using block adjacency graph for page segmentation and region identification; this has relevance to extraction of destination address block from postal mail image. Whichello and Hong (1996) have proposed a fast method for locating address blocks by dividing the image into clusters and classifying the clusters based on size, position and pixel density. Hase et al, (2000) have proposed a rule based method for extraction of strings from color document, which is useful for extraction of address strings from postal articles such as post cards, post envelopes and the like.
Madhvanath and Govindaraju (1999) have presented the case of local reference lines as opposed to a global reference lines for recognition of a line of text and have illustrated the success of the system that verifies street name phrases in a postal application. Wang et al, (1997) have proposed a method for detecting four reference lines for a cursive word at any skew angle. Gatos et al, (1997) have suggested a Hough transform based methodology for skew detection and text line position determination. Shivakumara et al, (2005) propose a new algorithm for accurate estimation of the skew angle using moments and boundary growing approach. Xiao and Leedham (2000) have presented a knowledge based approach to English cursive script segmentation using the background and character shape information. All these works refer to the preprocessing of text image for its further recognition.

Reading of mail addresses especially in the Indian context requires script/ language identification as the addresses are written in any of the different regional languages. Some works on script identification for Latin and Han languages are reported, whereas very few works are found for script / language identification in the Indian context. Spitz (1997) has proposed a system for identification of script and language content of a document. The system classifies the content into Han or Latin languages based on the concavities found. Later the Han based languages are recognized based on optical densities, whereas the Latin languages are classified based on character shape codes and word shape tokens. Tan (1998) has described a method for automatic script identification based on extraction of invariant texture features using Gabor filter technique. Hochberg et al, (1997) have proposed an automatic script identification technique, which clusters the templates of different scripts during training and the knowledge is used for classification. Patil and Subbareddy (2002) have presented a Probabilistic Neural Network (PNN) based method for script identification of a word. The methodology is used to identify English, Kannada and Devanagari scripts. Padma (2004) has proposed various methodologies using texture features to classify the images into different script classes for line and word images. The scripts considered for classification include Kannada, Tamil and other south Indian languages. There are very few reported works on script identification for Indian languages.
Character recognition is an inseparable part of postal automation. Character recognition has applications in other diverse areas such as document understanding, bank cheque reading, application form reading and number plate reading as well. Plenty of literature is available on character recognition. Some of the typical works are cited here. Shi et al, (2003) have presented a methodology for hand written basic shape recognition using non-linear active shape models and parameters from chamfer distance transform for Chinese language. Nagabhushan and Anami (2002) have proposed a knowledge based approach to recognition of Pitman Shorthand Language strokes. Arica and Fatos (2002) have discussed a Hidden Markov Model (HMM) for cursive handwriting recognition. These works bring out some of the various approaches adopted for character recognition.

Numeral recognition is a very important component of postal automation. The house and street numbers along with the PIN/ ZIP code are essential components of the destination address and need to be interpreted for dispatch and distribution sorting. Cheng and Yan (1998) have proposed a statistical approach using Fourier descriptors for English numeral recognition. Subbareddy and Nagabhushan (1998) have proposed a combination of neural network approaches for numeral recognition. Chi and Yan (1996) propose a fuzzy expert system approach to numeral recognition.

Some works have been reported for complete word/phrase recognition especially for postal address reading. Park (2002) has proposed a recursive approach based on interaction between flexible character classification and deductive decision making for recognition of handwritten word. Kim and Govindaraju (1998) have presented a neural network based approach for handwritten phrase recognition applied to street name images. El-Yacoubi et al, (2002) have proposed a statistical approach based on Hidden Markov Model (HMM) for location and recognition of phrases. These works have direct relevance to address interpretation stage of postal automation.

As the proposed work is to automate the processes of Indian postal system, the address reading techniques are to be developed for Indian languages. Very few works are reported in this context. Nagabhushan and Pai (1999) have proposed a region decomposition method by using bounding rectangles that grow in horizontal and vertical direction to

The text recognition techniques described above can be used/adapted to automate postal address reading; the interpreted addresses can be further used for mail sorting. The sorting of mail requires a lot of domain knowledge and there are very few reported works to tackle the knowledge based tasks such as address validation and mail delivery point (MDP) mapping.

Further the distribution of mail to the mail delivery points can be made efficient by generating optimal traversal path for the postman based on the mail to be distributed on a day-to-day basis. This generation of optimal mail distribution path is a challenging task, which needs assistance of appropriate computational methodology. The problem of optimal path generation for mail distribution is known in the literature as Chinese Postman Problem (CPP). Various cases of Chinese Postman Problem, such as Undirected Chinese Postman Problem (UCPP), Directed Chinese Postman Problem (DCPP) etc, exist and solutions to some cases of the problem have been described. Some of the problems such as UCPP, DCPP have polynomial time solutions \(O(n^3)\), where \(n\) is the number of nodes the traversal path has to visit under some constraints, whereas others such as Mixed Chinese Postman Problem (MCP) are NP-hard. Pearn (1994) discusses the solvable cases of the Chinese postman problem when k-postmen are considered. Pearn and Li (1994) present algorithms for solving the CPP when traversing in different directions of the road have different associated cost. Corberán et al, (2000) and Pearn and Wu (1994), present heuristic solutions to the Rural Postman Problem (RPP), where only a subset of edges of the graph are required to be visited whereas in CPP all the edges of the graph are to be visited. Though there are some heuristic solutions reported in literature for finding the optimal path for mail distribution, no work that integrates address interpretation and optimal path finding is reported.
The generation of optimal path for mail distribution requires an efficient representation for the locality served by the post offices. The simple graph representations used by the aforementioned solutions to CPP are not robust enough to represent the unstructured localities that are found in India. The Geographical Information System (GIS) technology has made available tools and concepts that can be used to represent such unstructured localities. Heywood et al, (2001) introduce various aspects of the Geographical Information Systems and provide an insight into the possible applications. Valerei and Aykut (2000) have proposed the use of Fuzzy objects in Geographical Information Systems (GIS). No reference to GIS representation for postal applications is found in literature.

The research papers listed in this section present the state of the art in areas relevant to postal automation. An insight into the motivation in taking up this research and computational strategies developed as part of the research are presented in sections 1.4 and 1.5.

1.4 MOTIVATION AND SYSTEM DESCRIPTION

The literature survey reveals that though fair amount of work has been carried out in building various sub systems of the postal automation system, no single reference to development of computational strategies for an Integrated Postal Automation System is found. This research is a step in that direction and aims at filling the gaps by developing techniques for integrating the various sub systems, such as address verification and validation using location specific knowledge, generation of optimal mail distribution routes for the postman from the identified mail delivery points etc. As the envisaged application is in the Indian postal system, techniques for reading of postal addresses written/ printed in English and Indian languages need to be developed.

The document processing tasks required for automating mail processing are found to be more efficient when location specific knowledge is employed. Also there is not much work reported in literature towards development of techniques suitable for reading Indian postal addresses. Hence there is a need to develop mail document processing strategies for
interpreting addresses in the Indian context taking the peculiarities of the Indian postal system into account.

The state of the art techniques for address reading described in section 1.3 are devised for reading the structured postal addresses of countries like United States of America (USA), United Kingdom (UK) etc [W-10]. But the postal addresses in countries like India are not structured [W-10] giving rise to a different set of problems. The postal addresses in India generally give a description of the geographical location of the delivery point of the addressee, for example, Near Playground, Behind CTO, Besides City Hospital etc. Further different people use different description of a mail delivery point; thus necessitating the step of mapping of address to mail delivery points after address interpretation, which is not essential for interpreting addresses in countries like USA, UK etc. Similarly the step of address validation more specifically verifying the correctness of PINCODE in relation to other parts of the address is necessary before forwarding it towards the destination.

Further the unstructured geographical layout of localities in India make direct use of arc routing solutions to optimal route finding difficult and necessitates development of GIS based techniques for finding optimal mail distribution paths/ routes. The optimal routes are to be found on a daily basis using the mail to be delivered on the given day and GIS information of the locality.

All the above factors motivated us to undertake research towards development of computational strategies for postal mail conveyance with the Indian postal system as a case study. The computational strategies developed for Indian system can be easily adapted to countries with a similar postal system. Further it is our view that postal automation systems are to be developed for a particular post office using the knowledge about the locality served by the post office. This is because the manual system works using the local knowledge, which makes most of the tasks efficacious. In consideration of these facts the issues in mail conveyance, with reference to Indian postal system addressed in this research are summarized below.
With a view to develop an integrated automation system for mail processing computational strategies are devised to deal with mail document analysis tasks, knowledge based tasks and optimization tasks. The document analysis tasks addressed in this research are identification of the script of postal address component, reading of PINCODE, automatic reading of words from a limited vocabulary, destination address block location and extraction of the characteristics of the component. The address validation, address component labeling and mapping the address to delivery point are the knowledge-based tasks addressed in this research. A geographical information system (GIS) representation for representing the beat of the postman is proposed. The GIS and soft computing techniques for finding the optimal and practically viable path for distribution of mail on a day-to day basis are also devised as part of the research.

The typical set of mail images found in the Indian Postal System that are to be handled by the proposed integrated postal automation system are depicted in Figure 1.4. The system model and computational strategies developed as part of this research are briefly described in section 1.5.

![Figure 1.4: A Typical Set of Mail Images](image-url)
1.5 PROPOSED SYSTEM MODEL

The Intelligent Integrated Automation System for mail processing requires computational strategies for automating the document analysis, knowledge based and optimization tasks to be employed in sorting, forwarding and distribution of mail. The proposed system for use in distribution or dispatch sorting post office comprises of the different modules depicted in Figure 1.5. The system comprises of seven modules as illustrated in the figure. The arrows bring out the relationship between the modules. The blocks labeled 1 to 7 represent function modules implementing the automation tasks. Modules 1 and 2 are simple knowledge base creation modules, whereas the other five modules address the various postal automation issues considering specifically faced problems in the Indian context.

Module 3 focuses on the various issues of document mail processing and address interpretation. This research work aims at developing computational strategies for tackling various issues of address interpretation in Indian context. The destination address block is located from the mail image by using positional rules. The different components are extracted and its script is identified using a newly proposed texture analysis approach. The other parameters of the address component such as the line number, number of words in a line etc are also extracted. A Fuzzy heuristic approach to recognition of PINCODE printed in Kannada script is developed. 7-segment projection based approaches to recognize PINCODE printed in Kannada and English scripts are introduced. Soft computing strategies using symbolic data analysis, fuzzy systems and neural networks for recognition of printed and hand written words in Kannada and English script as applied to postal address component reading are developed. In this research work, unless otherwise mentioned only English and Kannada scripts are considered for most of the tasks as address instances and mail images are taken from the mail found in the head post office of Bagalkot, a district place in Karnataka State of India. The local language of Bagalkot is Kannada and English is the common intermediate language in all the states of India.

Modules 4 and 5 comprise of computational strategies to emulate human expertise in mail sorting for forwarding to destination and delivery to the addressee and deal with the
knowledge-based tasks. A frame structured knowledge base approach to address validation and labeling of address components is developed as part of the research. More efficient approaches to address validation, address component labeling and mail delivery point mapping using symbolic data analysis and fuzzy techniques are also developed.

Modules 6 and 7 include tools for generation of optimal route for distributing the given day's mail. A Geographical Information System (GIS) representation for use in postal mail distribution application is proposed as part of this research. A graph theoretic approach to generation of a graph representing the roads on which the mail is to be delivered on a given day (mail distribution graph) is proposed. New methodologies using genetic algorithm and graph theoretic techniques are developed to generate the optimal and practically viable routes for distribution of mail. An adaptation of clustering methodology is devised to suggest various vehicle-parking positions for the postman, while distributing postal mail on a vehicle.

The system described above is to be developed for every sorting/ head post office in the country using local knowledge. The complete postal system can be made efficient by connecting the integrated postal automation systems in every sorting office by a high-speed computer network. The overall plan of such a system is described in chapter 12.

The computational strategies developed as part of this research make use of various technologies and tools such as image processing, neural networks, fuzzy systems, symbolic data analysis, genetic algorithms, cluster analysis and the like. A brief introduction to these technologies and their use in the current research work is presented in section 1.6.
### 1.6 TECHNOLOGIES EMPLOYED

The proposed Intelligent Integrated Automation System for Efficient Processing of Postal Mail requires automation of many tasks that make use of human intelligence. Automation of such tasks requires embedding intelligence into machines, by using machine learning techniques. In this research computational strategies for automating the sub tasks of mail
conveyance are developed by employing various technologies such as image processing, pattern recognition, cluster analysis and soft computing techniques such as fuzzy systems, neural networks, genetic algorithms etc. The following subsections provide an overview of these technologies and comment on their use in this research.

### 1.6.1 Image Processing and Pattern Recognition

Image processing technology is at the heart of computer vision tasks that aim to emulate the human vision process (Sonka et al., 2001). In a broad sense, image processing is processing of information for which both the input and output are images. Most image processing techniques treat the image as a two-dimensional digital signal. Digital computers are used to process the image to either extract some part of the image or enhance it. Availability of large computing power has given a tremendous boost to image processing applications. Image processing technology includes tools for noise removal, image enhancement, image analysis, image registration, image compression etc.

Automation of postal services require image processing tools adapted to reading of destination address and recognition of postal stamps among other tasks. Document layout analysis, word, character and numeral recognition as described in literature survey employ image-processing techniques. The word/ character/ numeral recognition tasks require concepts drawn from pattern recognition along with image processing tools.

Pattern recognition is a branch of artificial intelligence with techniques to cognize and recognize, patterns present in different kinds of data. Alternatively, it can be defined as the act of taking in raw data and taking an action based on the category of data (Theodoridis and Koutroumbas, 2006). Pattern recognition aims at classifying data (patterns) based on either *apriori* knowledge or on statistical information (features) extracted from the patterns. The patterns to be classified are usually groups of measurements or observations defining points in an appropriate multidimensional space. Pattern recognition and image processing techniques have been applied to automated analysis of medical images, automatic inspection of parts on assembly line, automatic grading of plywood, steel and the like, character/ word recognition, text recognition, etc. (Gose et al., 1997)
In the current research work pattern recognition and image processing techniques have been employed for document analysis, script recognition, numeral recognition and word recognition. The features extracted from the mail image after processing it are used to classify and recognize the scripts/numerals/words using decision theoretic or soft computing approaches.

1.6.1.1 Cluster Analysis

Pattern recognition techniques are also used to group similar objects without knowing their classes. This process is called as clustering. A cluster is a collection of objects, which are "similar" between them and are "dissimilar" to the objects belonging to other clusters. Clustering can be considered as an unsupervised learning problem. Given a set of feature vectors sampled from some population, the process of analyzing the feature vectors to ascertain the number of distinct data subsets and their determination is cluster analysis. Cluster analysis has been applied in many fields such as, image analysis, grouping of demographic data from census, medical imaging, marketing applications, city planning etc (Gose et al, 1997).

There are different types of clustering algorithms reported in literature. Some of them are, Exclusive Clustering, Overlapping Clustering, Hierarchical Clustering and Probabilistic Clustering. The hierarchical clustering is the most popular clustering methodology. There are many clustering algorithms such nearest neighbor, ISODATA, DBSCAN, BIRCH etc for grouping individual samples. (Gose et al, 1997) In this research work a modified version of GDBSCAN cluster analysis algorithm is adapted to group the mail delivery points in close proximity.

1.6.2 Soft Computing

Soft computing differs from conventional or hard computing in that, unlike hard computing, it is tolerant to imprecision, uncertainty, partial truth, and approximation. In effect, soft computing tries to emulate the working of human mind. There are many soft computing technologies used to build human like systems, prominent among them are fuzzy logic/systems, neural networks, genetic algorithms/ evolutionary computation. We
have also included symbolic data analysis a new tool for data analysis as a soft computing technology as it deals with imprecise descriptions using interval, qualitative and other forms of data (Bock and Diday, 2000). These soft computing technologies are complementary in nature and provide a more efficient system when used together. This research employs the combination of soft computing technologies to emulate human like activities in postal mail conveyance/processing. A brief introduction to these soft computing technologies is provided in the following sub sections.

1.6.2.1 Fuzzy Systems

Fuzzy systems are based on fuzzy logic as opposed to the conventional predicate logic. Fuzzy logic allows modeling of approximate and fuzzy concepts. Fuzzy logic proposed by Zadeh in 1965 provides a mechanism to model the vagueness in an idea (Kosko, 1997). For example if the room temperature on a given day is 14°C, then we may describe the temperature as cold or very cold depending on our perception. Such concepts are modeled in fuzzy systems using linguistic variables and membership functions (Kosko, 1997). Many methods to model various systems from machines to social systems are found in literature. The output of the fuzzy system is obtained by employing application specific defuzzification methodologies using fuzzy rules.

Fuzzy systems have found applications in, control systems for machines and industries, expert systems, machine learning, decision-making, classification and recognition etc. The use of fuzzy concepts in machine learning and building human like systems is on the increase. In the current research work fuzzy systems have been employed in numeral recognition for reading PINCODE printed in Kannada, labeling of address components and mapping the postal address to a mail delivery point. The ambiguity one faces in numeral recognition is well known, sometimes assigning labels to address components is also ambiguous as a component may seem to be a road or a house number, further a mail address may refer to more than one delivery point. In all these case the fuzzy techniques proposed in this work help in resolving the ambiguity.
1.6.2.2 Neural Networks

A neural network is a computing paradigm that is modeled after the neuron structures of the human brain. It consists of interconnected processing elements called neurons that work together to produce the required output. The neural network also called artificial neural network is an adaptive system that changes its structure based on external or internal information that flows through the network. Neural networks process information in parallel. The power of the neural networks comes from the algorithms designed to alter the strength/weights of the connections in the network to produce a desired signal flow. The interest in neural networks for various applications started with the description of a simple two state binary threshold type of neuron that has both excitatory and inhibitory inputs (Patterson, 1996). Though the early neural networks were designed based on the biological neural networks, the current day neural network implementations are based on statistical and signal processing concepts.

Neural Networks have emerged as a powerful pattern recognition technique to emulate humanlike behavior. Like other pattern recognition techniques the neural networks act on data by detecting some kind of underlying organization. For example scanned characters can be classified as different letters of the alphabet by detecting how they resemble each other. The networks can recognize spatial, temporal and other relationships and can perform such tasks as classification, prediction and function estimation. (Hammerstrom, 1993).

The neural networks infer solutions from data without prior knowledge about the regularities in the data, they learn the regularities empirically, that is they learn about the object from the data describing it. There are different types of neural networks available, some of them are feed forward networks, back propagation networks, radial basis neural networks, Kohonen self organizing maps, learning vector quantization networks etc. The different networks have been devised for different applications. In the current research work learning vector quantization (LVQ) neural network has been employed for recognition of the first character segmented from a cursive word using zonal features. The
abstraction and adaptation property of LVQ networks helps in overcoming errors in segmentation of first character.

1.6.2.3 Genetic Algorithms

Genetic algorithm is a soft computing tool that facilitates emulation of human intelligence in a machine. John Holland at the University of Michigan, USA first invented the genetic Algorithm (GA) in 1975 and subsequently it has been made widely popular by David Goldberg at the University of Illinois, USA (Mazumdar and Rudnick, 1999). The GA and its many variants are computational procedures that mimic the natural process of evolution and are also referred to evolutionary computational strategies. Genetic algorithms are search algorithms based on the mechanics of natural selection and natural genetics. They combine several of the fittest among string structures with a structured yet randomized information exchange to form a search algorithm emulating human behavior (Goldberg, 2000).

Genetic algorithms work by evolving a population of individuals over a number of generations. Genetic algorithms use two basic processes for evolution namely Inheritance and Competition. Inheritance refers to passing of features from one generation to the next by use of Crossover and Mutation operators. The operators are application specific and are defined for every application along with the representation of individuals. The crossover operator refers to the process of generating a new set of individuals from parents, whereas the mutation operator refers to the process of acquired changes in an individual. The process of Competition or Survival of the Fittest refers to weeding out unfit individuals from the population. A fitness value is assigned to each individual in the population, depending on the application and is used for selecting fit individuals. There are two types of genetic algorithms simple genetic algorithm and steady state genetic algorithm; they differ in the method employed for evolution of new generation. In simple GA the new generation completely replaces the old generation and the fittest individual may occur in any generation but not necessarily in the final generation. The new generation in steady state GA is obtained by replacing only weakest individuals in the previous generation and the fittest individual will be available in the final generation. Both the types of GA’s use
application specific terminating conditions. The performance of both types of genetic algorithms is satisfactory. Choice is dependant on the application. Since GA’s are heuristic procedures they are not guaranteed to find the optimum, but often they are able to find very good solutions for a wide range of problems. Goldberg (2000) states that, *Convergence to the best is not an issue in business or in most walks of life we are only concerned with doing better relative to others, Hence the most important goal of optimization is improvement.*

Hence genetic algorithms are preferred for optimization tasks in variety of applications. Genetic algorithms find applications in optimization, search, decision making and machine learning. In this research work genetic algorithms are employed for finding the maximal matching of odd degree vertices, the optimal local tour in a cluster of mail delivery points, etc. The chromosome representation, cross over and mutation operators are defined for the said applications.

### 1.6.2.4 Symbolic Data Analysis

Symbolic objects are extensions of classical data types and they are defined by logical conjunction of events linking values and variables in which the variables can take one or more values and all symbolic objects need not be defined on the same variables. Based on the complexity the symbolic objects can be of three different types namely assertion, hoard and synthetic. An assertion object is a conjunction of events pertaining to a given object. An event is a pair which links feature variables and feature values. A hoard object is a collection of one or more assertion objects, whereas a synthetic object is a collection of one or more hoard objects (Gowda, 2004). Diday (2004) says a symbolic object and its extent can model a concept of the real world.

Diday (2004); Bock and Diday (2000) describe the extension of standard exploratory data analysis to symbolic object analysis. The inputs to symbolic data analysis are symbolic data object and a database/ knowledge base referred to as symbolic data table (similar to a table in a relational data base). The processing algorithm is application dependent. Gowda (2004) proposes many similarity and dissimilarity measures that can be employed in symbolic data analysis. A clustering algorithm using similarity measure is also described
by (Gowda, 2004). The output of the processing algorithm is either a description of the
class or an individual object depending on the application.

The symbolic data analysis has been extensively used in this research. A mail delivery
point can be described using more than one address; hence symbolic representation is
better suited to represent a postal address. A new symbolic object representation of postal
address and various symbolic knowledge bases for address validation, address component
labeling and mail delivery point mapping are employed in the research. A fuzzy symbolic
methodology for address component labeling and mail delivery point mapping is also
developed. A symbolic representation of a character and a word and their use in character/
word recognition using fuzzy symbolic approach are proposed.

Various computational strategies are devised in this research using the technologies
described in this section. The contributions of the research are listed in section 1.7.

1.7 CONTRIBUTIONS OF THE RESEARCH

The research for development of computational strategies to automate postal service in the
Indian context has lead to the solution of various technical issues. The strategies developed
can find application in postal automation and other related application areas as well. The
research is to provide integration of various aspects related to postal automation. Different
aspects of postal automation have warranted research in different directions. For this
reason the contributions are classified into three parts to highlight the different aspects of
the research namely, Document Image Analysis/Computer Vision Tasks, Knowledge based
tasks and Optimization tasks.

Contributions towards Document Image Analysis/ Computer Vision Tasks (PART I)

- A fuzzy statistical approach for recognition of Kannada numerals and thus the
  PINCODE printed in Kannada.

- Texture based approach for identification of scripts of the PINCODE image.
Symbolic similarity analysis approach for recognition of words printed in English and Kannada after identifying the script of the word image using geometrical features.

A symbolic similarity measure for recognition of printed words.

Soft computing strategy for recognition of handwritten cursive words, employing LVQ neural network and fuzzy symbolic similarity analysis.

A symbolic similarity measure for recognition of a segmented character of a handwritten word.

A symbolic similarity measure to identify a handwritten word from a limited lexicon using a segmented part of the word.

Strategies for outer contour extraction, segmentation of handwritten word image, skew correction, slant correction, extraction of ascender/descender features.

Contributions towards Knowledge based Tasks (PART II)

A frame supported knowledge base approach for postal address validation.

A frame supported knowledge base approach for address component labeling.

A symbolic representation for postal address and symbolic data analysis methodology for postal address validation.

A fuzzy symbolic methodology for extraction and labeling of postal address components.

A similarity measure for labeling of address components.

A soft computing strategy for mapping mail address to mail delivery point.

A similarity measure for mapping address to mail delivery point.
Contributions towards Optimization Tasks (PART III)

- A geographical information system (GIS) representation of the beat of the postman highlighting the mail delivery points and other entities useful in computing the optimal path for distribution of mail.

- Spatial data structure for representing beat of the postman useful in generating a connected graph from an unconnected graph representing the roads on which mail is to be distributed.

- A new routing problem to represent the postal mail distribution optimization.

- An overall solution strategy for postal mail distribution optimization using spatial data structure, genetic algorithms and graph transformations.

- A steady state genetic algorithm for optimal matching of odd degree vertices in a graph.

- An overall solution strategy to find a practically optimal path for distribution of mail by a postman using traverse-park-distribute strategy when the postman uses two modes of transport (vehicle and foot) for distribution of mail.

- A modified GDBSCAN algorithm to identify clusters of active mail delivery points (mail delivery points that receive mail) and locating the parking spot for the vehicle of the postman.

- A steady state genetic algorithm for finding an optimal local tour (Hamiltonian Path) for distribution of mail in a cluster of mail delivery points.

These new techniques are described in detail in this thesis; Section 1.8 gives the organization of the thesis.
1.8 ORGANIZATION OF THE THESIS

As presented in the previous section, the contents are classified into three parts projecting the different aspects of postal automation and in the sequel enabling proper presentation of the research highlighting all the different aspects. Accordingly this thesis is organized into three parts and consists of twelve chapters.

Chapter 1 gives a detailed introduction to the research work and the thesis.

Part I describes the document analysis and address reading techniques developed as part of the research. The Part I of the thesis is divided into four chapters, chapters 2 to 5. Chapter 2 presents computational strategies for identification of the address block and further identification of the script of the PINCODE image. Chapter 3 describes a computational mechanism akin to human approach for fast reading of word printed in Kannada and English from a limited lexicon using symbolic similarity analysis of the geometrical features representing the word image. Chapter 4 presents human inspired approach for reading of cursive words using soft computing methodologies. Chapter 5 presents some new methodologies for recognizing the PINCODE images printed in Kannada and English.

Part II presents the knowledge-based techniques for postal address validation and mapping mail address to delivery points. Part II is further divided into 3 chapters namely chapters 6, 7 and 8. Chapter 6 describes a knowledge-based approach using frames for labeling of the components of postal address and validation of the address for dispatch sorting. Chapter 7 discusses the symbolic object representation of the postal address and the soft computing methodologies for address component labeling and address validation. Chapter 8 presents the soft computing strategies for mapping postal addresses to mail delivery points.

Part III describes GIS, graph transformation and soft computing based techniques developed for finding the optimal and practically viable routes for distribution of postal mail. Part III is divided into 3 chapters, namely chapters 9, 10 and 11. Chapter 9 discusses the optimization issues in distribution of postal mail and the conventional approach for finding the optimal path for distribution of postal mail. The optimization of the mail distribution is presented in a new paradigm and GIS representation for the relevant entities.
is proposed. Chapter 10 describes a computational strategy for finding the optimal path and schedule for distribution of mail using GIS representation genetic algorithm and graph transformation techniques. Chapter 11 presents a GLOCAL approach using GIS representation, cluster analysis, graph transformation and genetic algorithms for finding a practically viable path for distribution of postal mail.

The overall scheme for integrated postal automation system, conclusions drawn from the research and future avenues are presented in chapter 12.
APPENDIX TO CHAPTER 1

THE INDIAN POSTAL SYSTEM

AC1.0 INTRODUCTION

The Indian postal system was established in the year 1764 and is one of the oldest postal systems of the world. The Indian Postal System has the largest postal network in the world, servicing about 1,55,516 post offices (as on 31.3.2005). Department of Posts, Government of India, manages the postal service in India. The department of posts functions under the Ministry of Communications and Information Technology. The Postal Services Board is the apex management body of the department, comprising the chairman and three members. The postal service is offered under the brand name of India Post.

The vision of India Post is “to be a socially committed, technology driven, professionally managed & forward looking organization” (refer to www.indiapost.gov.in for further details). The department of posts provides universal access to basic postal services in the country at affordable prices. The most important postal service is the conveyance of mail from sender to receiver. The department of posts transports the written word messages (mail) either using standard postal stationary such as post card, inland letter, postal envelope or non-standard envelopes. For providing postal services, the whole country has been divided into twenty-two postal circles. The postal circles are enlisted in Table AC1.1. Each circle is associated with a state except a few which cater to more than one state. Over and above the twenty-two postal circles, the army base circle caters to the communication needs of the armed forces.

A Principal Chief Postmaster General or a Chief Postmaster General heads the postal circle. Each circle is further divided into regions comprising of groups of field units, called divisions. There are two types of divisions namely Postal Division and RMS (Railway Mail Service) Division. The postal division is assigned the task of delivery and collection of mail, whereas the RMS division deals with sorting and conveyance of mail.
Other supporting units like Stamp Depots, Store Depots and Mail Motor Service assist the postal system.

**Table AC1.1: List of Postal Circles in India**

<table>
<thead>
<tr>
<th>SI No</th>
<th>Postal Circle</th>
<th>SI No</th>
<th>Postal Circle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Andhra Pradesh</td>
<td>2</td>
<td>Assam</td>
</tr>
<tr>
<td>3</td>
<td>Bihar</td>
<td>4</td>
<td>Chattisgrah</td>
</tr>
<tr>
<td>5</td>
<td>Delhi</td>
<td>6</td>
<td>Gujarat (Also covers the areas of Daman &amp; Diu, Dadra and Nagar Haveli)</td>
</tr>
<tr>
<td>7</td>
<td>Haryana</td>
<td>8</td>
<td>Himachal Pradesh</td>
</tr>
<tr>
<td>9</td>
<td>Jammu &amp; Kashmir</td>
<td>10</td>
<td>Jharkhand</td>
</tr>
<tr>
<td>11</td>
<td>Karnataka</td>
<td>12</td>
<td>Kerala (Also includes the Union Territory of Lakshadweep)</td>
</tr>
<tr>
<td>13</td>
<td>Madhya Pradesh</td>
<td>14</td>
<td>Maharashtra (Also includes Goa)</td>
</tr>
<tr>
<td>15</td>
<td>North East (consists of Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland and Tripura)</td>
<td>16</td>
<td>Orissa</td>
</tr>
<tr>
<td>17</td>
<td>Punjab (includes the Union Territory of Chandigarh)</td>
<td>18</td>
<td>Rajasthan</td>
</tr>
<tr>
<td>19</td>
<td>Tamilnadu (Also includes includes Pondicherry)</td>
<td>20</td>
<td>UttarPradesh</td>
</tr>
<tr>
<td>21</td>
<td>Uttarakhand</td>
<td>22</td>
<td>West Bengal (includes Sikkim and the Union Territory of Andaman &amp; Nicobar Islands)</td>
</tr>
<tr>
<td>23</td>
<td>Army Postal Service (Base Circle)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*India Post* offers postal service through a network of post offices spread across the country. Post offices in the country are categorized as Head, Sub and Branch Post Office based on the functionality assigned to them. Gramin Dak Sewa Post Offices are located in rural areas. The postal network is being expanded in rural areas of the country by opening Extra Departmental Post Offices. The Extra Departmental Post Office (EDPO) functions on part-time basis for minimum of three hours and a maximum of five hours. Part-time locally recruited employees, who are paid an allowance for the services they render, manage the EDPO’s.

Post offices in urban and rural areas provide a range of basic postal services. With time the postal system has diversified into financial sector for achieving self-sufficiency. The diversification into financial services sector was a conscious decision to take advantage of the core strength of the postal system, which is “the outreach of the network, a well
established accounting system and intimate knowledge of local conditions even in the remotest corners of the country".

The Indian Postal System currently provides a number of services which can be broadly classified into four categories: Communication services (Letters, Post Cards, etc.), Transportation services (Parcel, Logistics Post), Financial services (Savings Bank, Money Order, International Money Transfer Service, Public-Private Partnerships for extending financial service outreach through the post office network, Postal Life Insurance) and Premium Value Added Services (Like Speed Post, Business Post, Retail Post).

The value added postal services offered by India Post in recent years include Speed Post, Business Post, Express Parcel Post, Greetings Post, Data Post, Speed Post, Passport Service, Bill Mail Post, e-Post and e-Bill Post etc. In the area of financial services, new services introduced include facilities for International Money Transfer, Electronic Fund Transfer, Electronic Clearance Services, Warrant Payment, Sale of Mutual Funds and Bonds etc.

The postal department is working towards improving the quality of postal service its core business area. This appendix describes the postal operations in India, the efforts of *India Post* to computerize its operations and the challenges faced by *India Post*.

The remaining part of this appendix is divided into five sections. The section AC1.1 describes the postal operations of India Post. Section AC1.2 discusses the PINCODE (Postal Index Number code), the postal code introduced by India Post to assist in efficient sorting of Postal Mail. The section AC 1.3 describes the efforts of the department of posts for adapting to advances in technology and computerization. The section AC1.4 gives a brief introduction to MEGHDOOT software used by India Post. The section 1.5 summarizes the challenges faced by India Posts.
AC1.1 POSTAL OPERATIONS

Postal operations at the post office encompass the entire gamut of the basic postal services, which include the following activities.

- Conveyance/transport of mail/parcel
- Sale of stamps and stationery such as post card, inland letter, postal envelope etc
- Booking of registered articles
- Booking of insured articles
- Booking of value payable articles
- Remittance of money through money orders and postal orders
- Booking of parcels

Mail processing or mail conveyance is the core activity of the India Post. Mail is collected from 5,84,006 letter boxes in the country. This is processed by a network of 460 Railway Mail Service (RMS) offices and is transported by road, rail and airlines all over the country. The Indian postal system has the largest postal network in the world, servicing about 1,55,516 Post Offices (as on 31.3.2005), out of which 1,25,148 are Branch Post Offices in rural areas. On an average, a Post Office serves an area of 21.13 sq kms and a population of about 6,615. The total amount of registered and unregistered mail traffic in the year 2004-05 was about 7360(215-registered) million mail pieces.

The collection of mail from letterboxes and delivery of mail to the addressee is done by postmen employed by India Post. The most important subtask of postal service is sorting of mail. The mail is sorted at origin, intermediate and delivery post offices. The mail is sorted at the delivery post office for distribution of the mail to the addressee. This sorting is done to segregate the mail into different beats (contiguous part of the area which is assigned to the postman for delivery of mail).

The mail is sorted at the origin and intermediate post offices for forwarding it towards the destination. This type of sorting is referred to as dispatch sorting. The dispatch sorting procedure adapted by India Post is described in the following. Every sorting office has one
Indian Postal System

or more set of people called a sorting set. The head-sorting assistant leads the sorting set. The sorting set sorts the outgoing mail and sends them either to an intermediate sorting office for further sorting or to a delivery post offices for delivery. The overall guiding principle employed in sorting of mail is to forward the mail so as to reach the destination in the least possible time and no sorting office keeps the mail without forwarding. Using this principle the sorting is done by forwarding a letter to the destination place if a direct bag is sent to the place, otherwise the letter is forwarded to a place nearby to the destination place. If a letter is destined outside the state/ postal circle and there is no direct bag sent to the place and it is not nearby to the places to which direct bags are sent then the letter is forwarded to the circle head quarters. The circle head quarter then forwards it to the destination circle. The sorters knowledge of geography and the place names plays a significant role in proper sorting of mail. There are many places with same names; in such cases the sorters generally make use of other related information for proper sorting and forwarding of mail. If the destination address does not contain a clear indication of the destination so that it can be sorted and forwarded, it is sent to dead letter office at the circle head quarters for further resolution or storage.

To facilitate speedier sorting of mail and sorting automation, India Post has introduced a six digit postal code called the Postal Index Number (PIN). Section AC1.2 gives a brief description of the PINCODE.

**AC1.2 PINCODE**

The Postal Index Number called PIN in short is an all-numeric postal code introduced by Department of Posts on 15\textsuperscript{th} August 1972. The PIN also called PINCODE is a unique six-digit number assigned to every delivery post office. The branch post offices use the number allotted to their parent accounts post office. In a country like India, which has many languages and scripts, the introduction of the numeric code has eased the problem of mail sorting. Further there are several different towns in the country having the same name; these towns can be easily distinguished using the unique PINCODE assigned to them. The addition of the PIN code after the address enables the sorting personnel to identify the destination correctly, thereby eliminating chances of mis-sending and the
consequent delay in the conveyance of the article. India Post recommends that the PIN should be the last item of information in the address, preferably on a line by itself. If this is not possible, it may be added after the last item of the address.

The digits of the PINCODE identify the geographical position of the distribution post office, which is to deliver the mail. The country has been divided into eight different postal zones and the first digit identifies the zone to which the delivery post office belongs. The digits identifying the zone are enlisted in Table AC1.2. The first two digits from the left identify the postal circle of the destination address and is depicted in the PINCODE map of Figure AC1.1. The first three digits of the PINCODE taken together indicate the sorting office to which the distribution post office is affiliated. The last three digits of the PINCODE identify the delivery post office.

**Table AC1.2 : The first digit of PIN**

<table>
<thead>
<tr>
<th>First Digit</th>
<th>Region</th>
<th>States Covered</th>
<th>First Digit</th>
<th>Region</th>
<th>States Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Northern</td>
<td>Delhi, Haryana, Punjab, Himachal Pradesh, Jammu &amp; Kashmir</td>
<td>5</td>
<td>Southern</td>
<td>Andhra Pradesh and Karnataka</td>
</tr>
<tr>
<td>2</td>
<td>Northern</td>
<td>Uttar Pradesh and Uttarakhand</td>
<td>6</td>
<td>Southern</td>
<td>Kerala and Tamil Nadu</td>
</tr>
<tr>
<td>3</td>
<td>Western</td>
<td>Rajasthan and Gujarat</td>
<td>7</td>
<td>Eastern</td>
<td>West Bengal, Orissa and North Eastern</td>
</tr>
<tr>
<td>4</td>
<td>Western</td>
<td>Chattisgarh, Maharashtra and Madhya Pradesh</td>
<td>8</td>
<td>Eastern</td>
<td>Bihar &amp; Jharkand</td>
</tr>
</tbody>
</table>

*Note: No valid PINCODE has '9' as the first digit*

The department of posts maintains a website through which the PIN code of any delivery post office can be found. The use of PINCODE in the mail makes the task of mail sorting easy, but though *India Post* is trying to popularize the use of PINCODE, it is not well known and still people write destination address without PINCODE or some times partial or erroneous PINCODE. Along with the introduction and promotion of the use of postal code, India Post is adopting modern technology for improving the efficiency of the postal service.
AC1.3 COMPUTERIZATION OF POSTAL SERVICES

In order to expedite transmission, processing and delivery of mail, India Post has been pursuing a policy of computerization and modernization of the mail processing, transmission and delivery in post offices. The modernization of post offices aims at providing better equipment and improved work environment in the post offices. Sorting of

Figure AC1.1: PINCODE Map of India (from www.indiapost.gov.in)
accountable articles like Registered Letters, Insured Articles etc, has been computerized for speedy and safe handling of articles and for more expeditious response in case of enquiries, using bar codes. Internet based Track and Trace System has been implemented for Speed Post and Express Parcel Post wherein a customer can find out the status of his consignment online through the Internet by logging on to the website www.indiapost.gov.in.

At present two Automated Mail Processing Centers (AMPCs) are functioning at Mumbai and Chennai for automatic sorting of ordinary postal mail. The India Post has initiated action for setting up of two more Automatic Mail Processing Centres (AMPCs) at Delhi and Kolkata. These AMPCs use imported postal mechanization equipment to automate sorting of ordinary mail. The performance of mechanized sorting is not satisfactory as the techniques devised for reading the address are not suitable for the Indian context and there are no tools to emulate the human skill in sorting of mail. Hence there is a need to develop postal automation strategies suitable for the Indian postal system.

Further across the counter postal services were being provided at manually operated counters of the post office. To improve the efficiency of the services offered, the operations at the counter are now being progressively computerized using software called MEGHDHOOT. MEGHDHOOT provides a greater range services to the customer from a single window thus contributing to the services being more responsive and error free. The department of posts has a target to computerize all 839 Head Post Offices, 6861 large Sub-Post Offices, 22 Postal Accounts Offices and 264 Administrative Offices during the Tenth Five Year Plan that ends in 2007. The postal department is also working towards networking of post offices by setting up a National Data Centre in National Informatics Centre (NIC).

**AC1.4 MEGHDHOOT SOFTWARE**

MEGHDHOOT is a software developed by the software development team of India Post to computerize across the counter/ point of sale services. All the operations of the counter except savings bank transactions can be carried out through this package. The software is
capable of handling all transactions that are performed at the post office counters and are listed as under,

- Booking of all categories of accountable articles of inland mail like registered letters/parcels/packets, insured letters/parcels, value payable (VP) articles etc.
- Undertaking financial transactions like booking of various categories of money orders, accepting of postal life insurance premium, telephone bills, miscellaneous receipts, sale and payment of Indian Postal Orders etc.
- Booking of domestic and foreign articles for speed post.
- Booking of all categories of accountable articles of foreign mail.
- Attending to PIN code enquiry etc.

The Postman module of MEGHDHOOT software, which aims at computerizing the delivery work at the post office, is capable of handling the following works.

- Receipt of registered/VP/insured letters/parcels, issue of articles to postmen, tracking delivery of the articles issued, delivering articles across the counter and maintaining accurate account of the articles handled.
- Issue of money orders received for payment to postmen, watching delivery, payment across the counter and maintaining account thereof
- Maintaining account of unregistered articles received for delivery that are unpaid/insufficiently paid, issuing these to postmen and watching delivery thereof
- Beat wise sorting of the accountable articles can be done if the locality database has been entered and the destination address is typed in (presorting). Otherwise manual sorting is to be done and the information is to be entered in the system (Post sorting)
- It maintains the mail returned/redirected information also. It also has interface with treasury module for MO information.
- Changes in beat allocation can be done manually.
The other modules of the software include, sub-account, treasurer, Speednet (for tracking speedpost articles), SpeedCCC (for customer care etc). More details can be found at http://www.ptcinfo.org. India Post has also customized this software for operations of Bhutan Post and Uzbekistan Post in the postal departments of the respective countries.

**AC 1.5 CHALLENGES FACED BY INDIA POST**

The advances in electronic communication technology are making postal mail conveyance outdated leading to steady fall in revenues for the postal department. Added to this the postal department is now facing stiff competition from a plethora of private courier service providers. The department of posts has responded by diversifying into financial services and is trying to make the best use of the available technology to provide efficient and value added service for mail conveyance.

The imported postal service mechanization equipments are not performing as expected and their use has not enhanced the efficiency of the postal service, this is because the techniques employed in these imported machines are not suitable in the Indian context. Hence there is a need to develop computational strategies for automation of the Indian Postal System. The problems associated with automation of Indian Postal services are enumerated in this research and novel computational strategies are devised and tested using postal data of Bagalkot, a district place in Karnataka State of India.
PART I

AUTOMATIC EXTRACTION AND READING OF POSTAL ADDRESS
OVERVIEW OF PART I

The postal mail conveyance can be made efficient by automating its subtasks. The most important subtask of postal mail conveyance that requires automation is the sorting of postal mail. The sorting of postal mail requires automatic reading of the destination address. The address reading entails identification of the destination postal address from the postal mail article, separation of different parts of the address, extraction of their characteristics and further recognizing the address for understanding the destination postal address. All of these are postal document image analysis tasks.

The Part I of the thesis deals with only a few postal document analysis issues in order to concentrate on other aspects of integrated postal automation described in Part II and Part III. Some new computational strategies for reading of addresses especially in the Indian context are presented in Part I. The premise of the new strategies presented for address recognition is that, the computational strategies for address reading are to be devised using the local knowledge at the post office where the address recognition is to be carried out, employing human like approaches.

The Part I of the thesis, is divided in to four chapters. Chapter 2 presents computational strategies for identification of destination postal address and separation of the address components based on their line of occurrence. A new methodology based on texture analysis is presented for identification of the script of the printed PINCODE images. Chapter 3 presents a new methodology based on symbolic data analysis for reading of address words printed in English and Kannada encountered in a particular post office. The methodology emulates the human word reading mechanism and identifies the word in a limited lexicon by processing the geometric features of the word image. Chapter 4 presents novel soft computing strategies for identifying the cursive words of an address in a limited vocabulary when reading addresses encountered in a particular post office. The cursive word reading methodologies use the geometrical features of the word image. New computational strategies for recognizing the PINCODE printed in Kannada using the texture features are presented in chapter 5. New 7-segment projection approaches for reading PINCODE printed in English and Kannada are introduced.