Aum Sri Sairam

# DESTINATION ADDRESS INTERPRETATION FOR AUTOMATING THE SORTING PROCESS OF INDIAN POSTAL SYSTEM

Guided by

Mrs. C. Vasanthanayaki, M.E.

P. Sri Rama Prasanna

S. Balaji

Thejavor Haralu Khezhie

B.E Thesis Presentation, April 2003, GCT Coimbatore

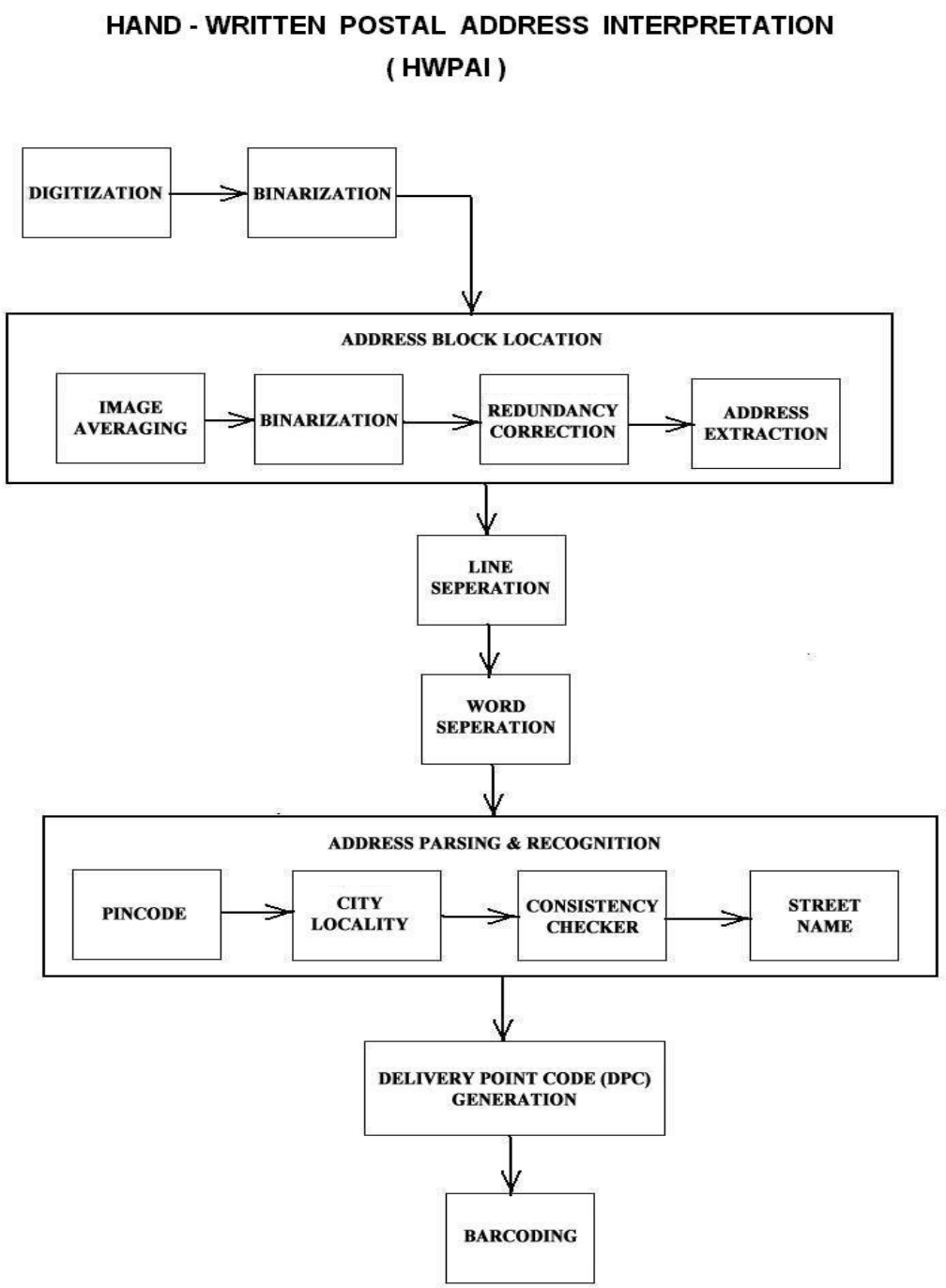
# Introduction

- → Present Postal System
- → Bottlenecks in existence
- → Need for improvement
- → What is this project for ?

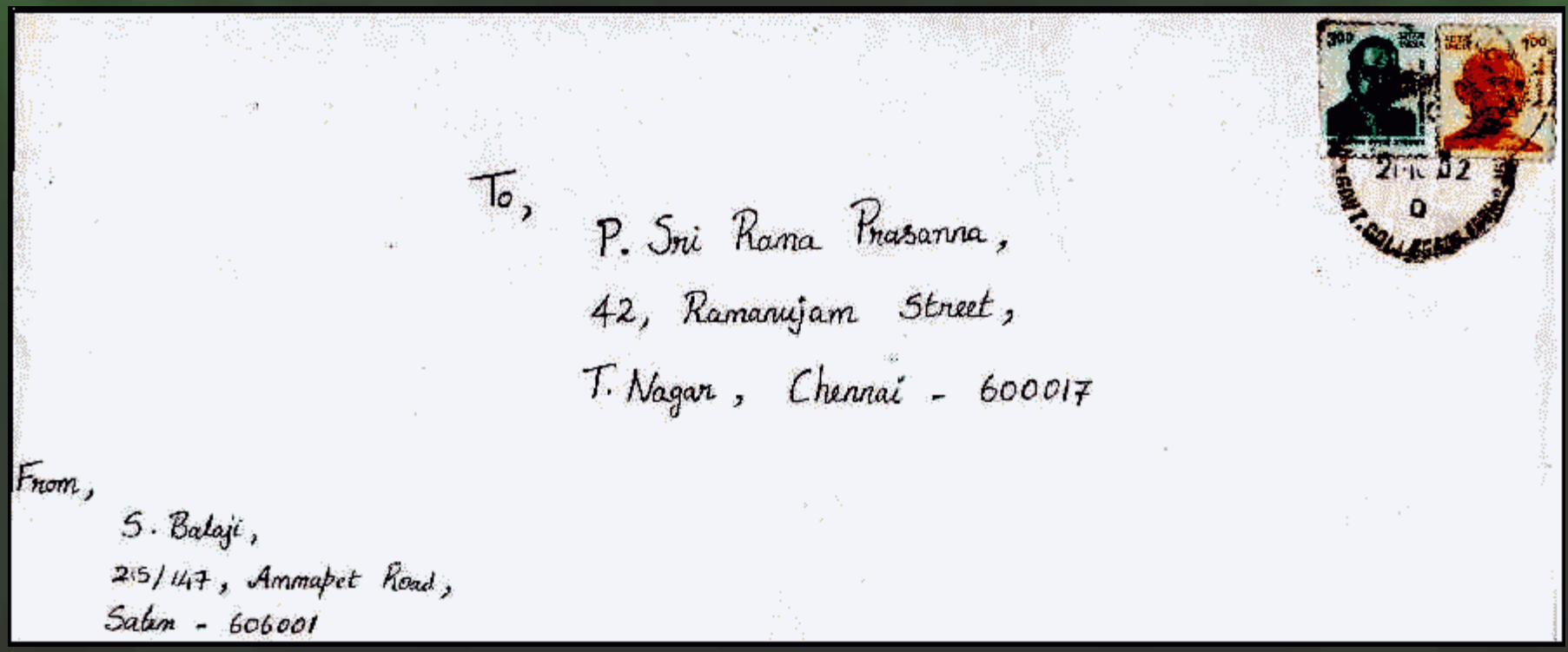
# Our *Project* in a nutshell

- → Digitization
- → Preprocessing
- → Address Block Location
- → Line and Word Separation
- → Address Parsing and Recognition
- → Delivery Point Code (DPC) generation
- → Bar-coding

# Block Diagram

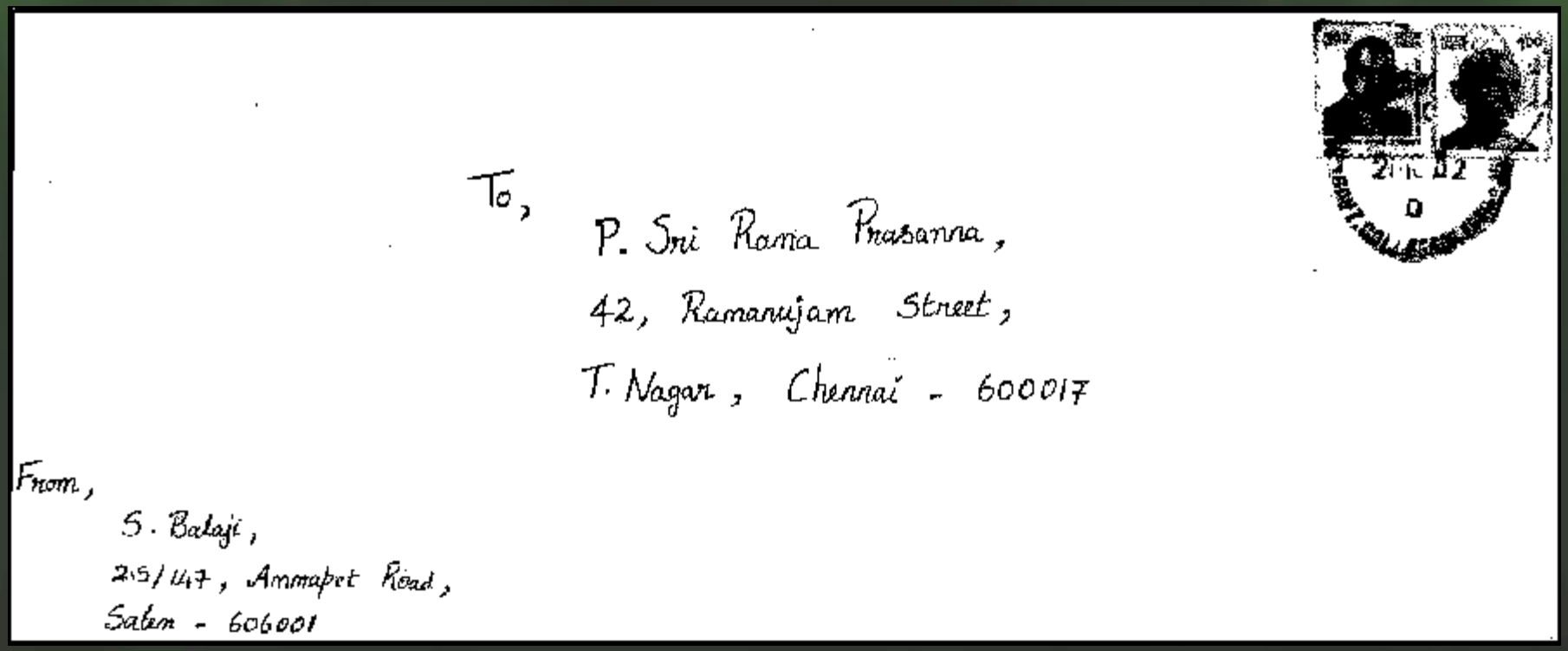


# Digitization



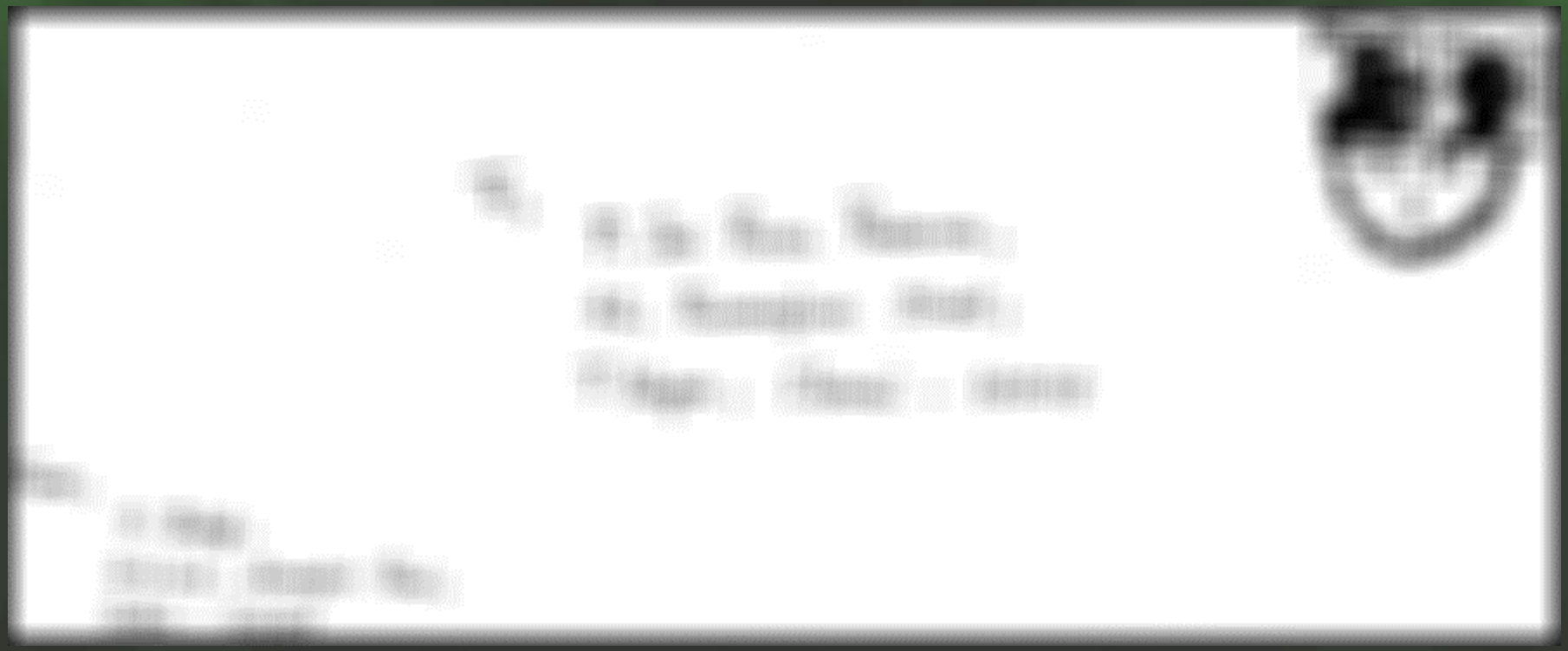

The envelope is **digitized** at 212 dots-per-inch (dpi), in 8-bit grayscale.

# Binarization



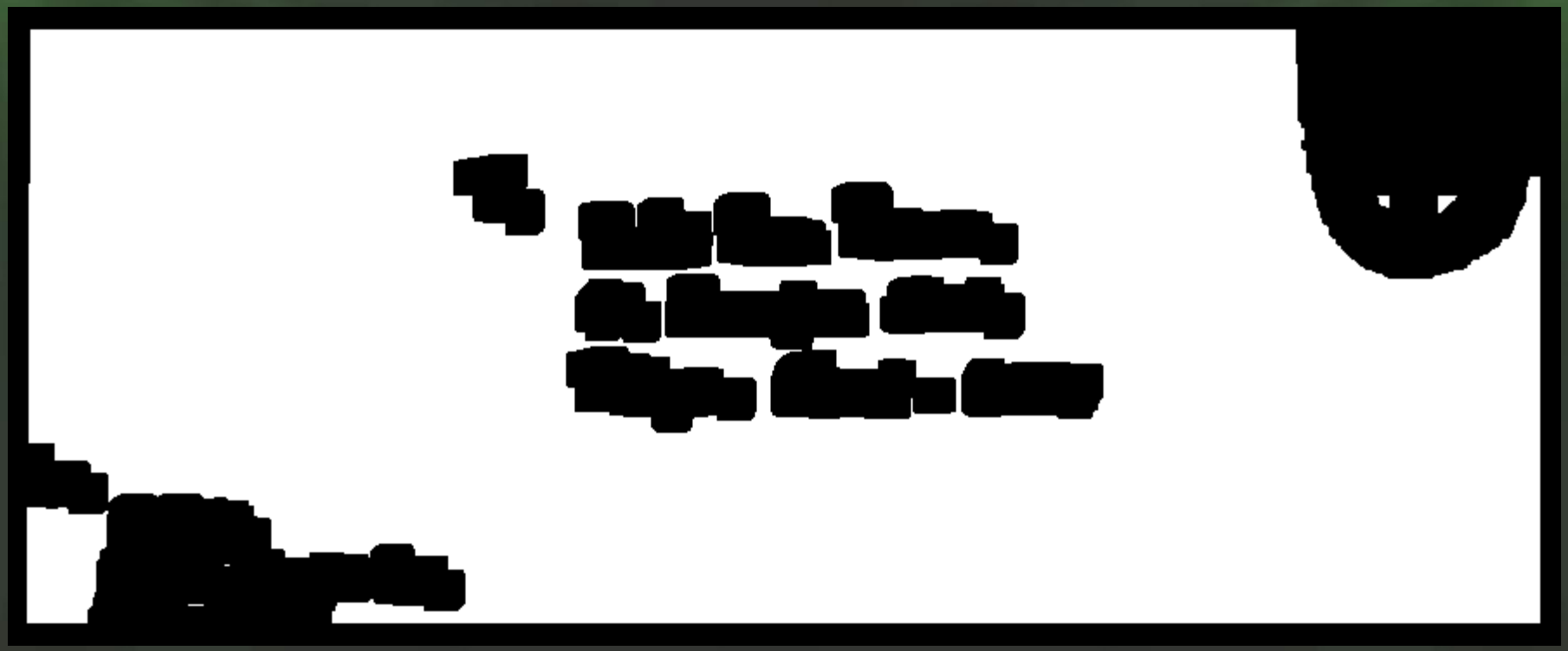

The digitized envelope is binarized using im2bw() function in MATLAB

# Image Averaging



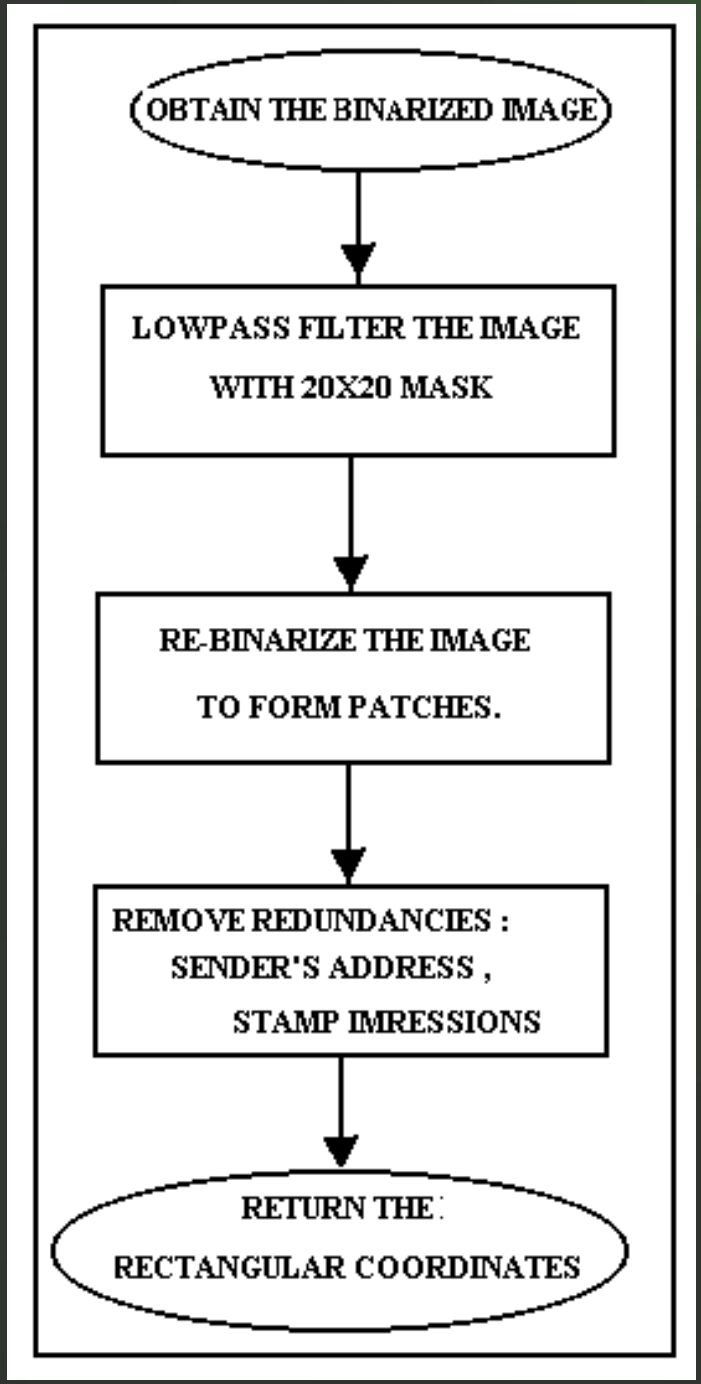
The binarized image is low-pass filtered with a 20 X 20 mask

# Re - Binarization



The filtered image is re-binarized using im2bw() function with a level 1.0

# Redundancy Correction Algorithm



# Redundancy Corrected Output

The redundant information in the corners are removed

# Address Extraction

To,

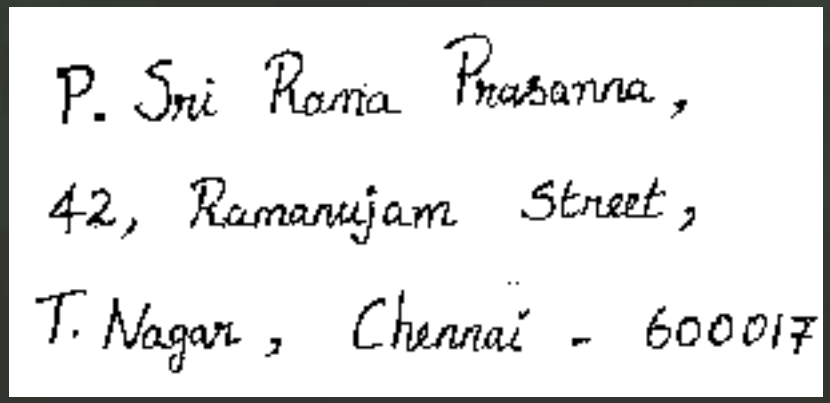
P. Sri Rama Prasanna,
42, Ramanujam Street,
T. Nagar, Chennai - 600017

From,

S. Balaji,
215/147, Ammapet Road,
Salem - 606001

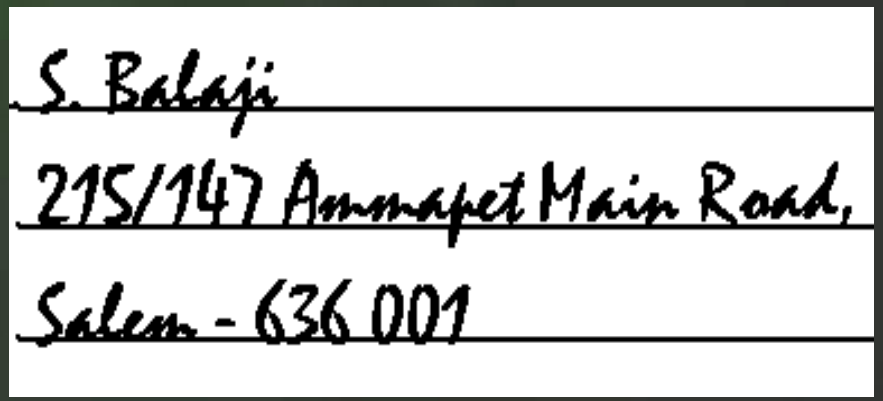
From the Co-ordinates obtained from ABL module, the address is extracted.

# Extracted Address




imcrop() function in MATLAB is used for extraction

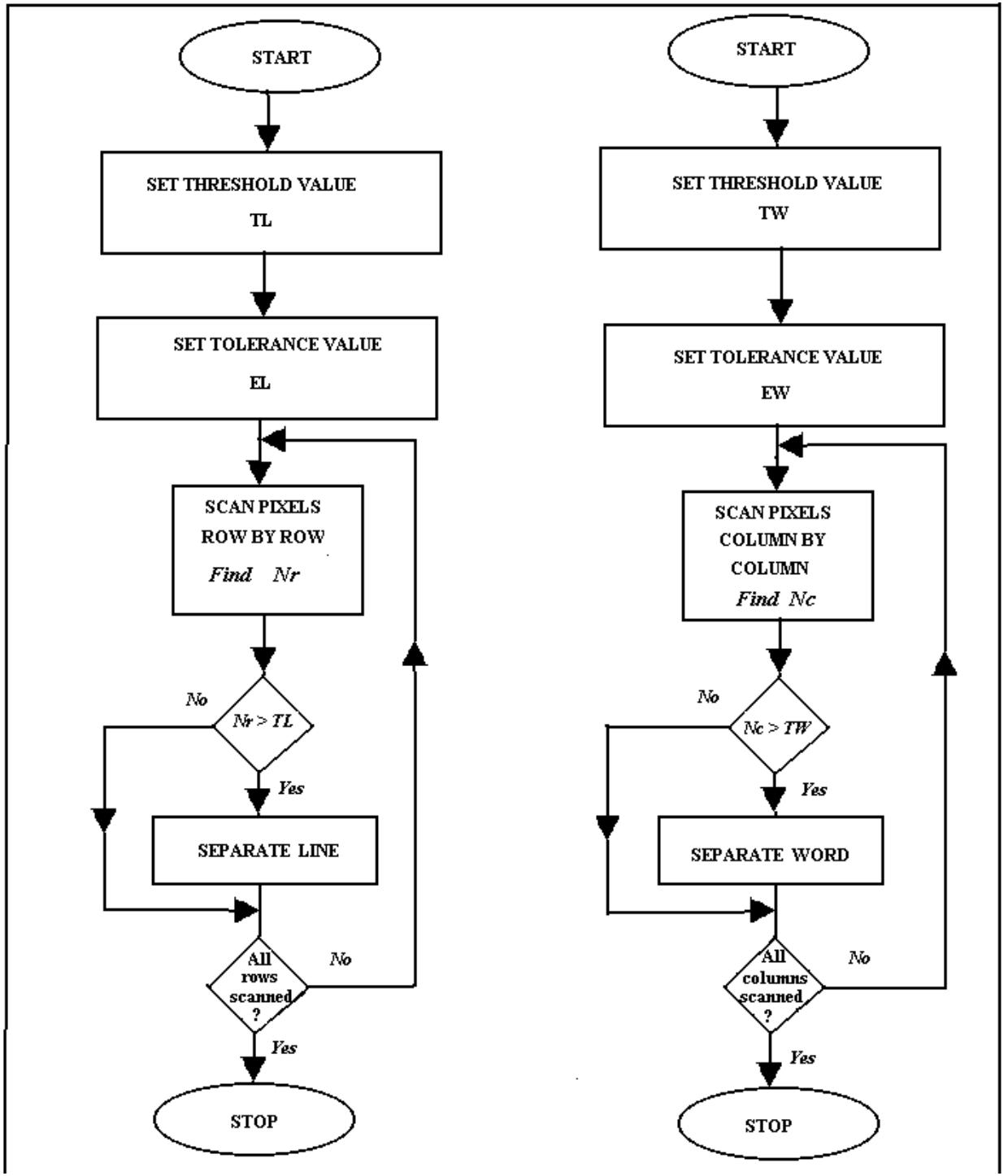
# Underline Removal





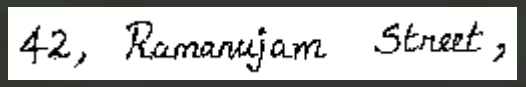
- The Underlines are removed by using Top Pixel Algorithm
- This prevents the possibility of destination address getting tampered.

Horizontal Scanning




Vertical Scanning

# Line Separation

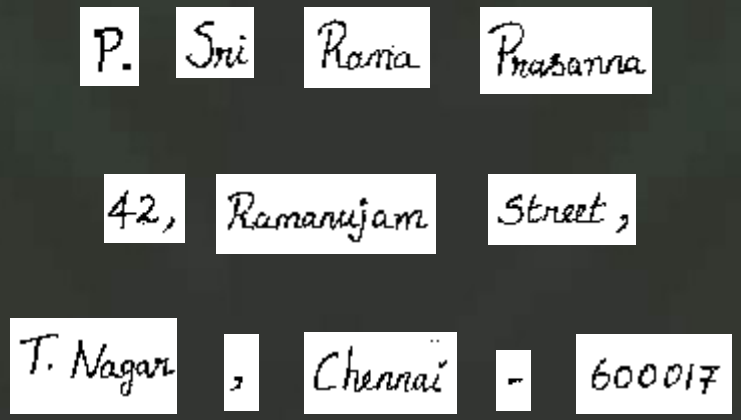





The individual lines are separated by specifying a threshold value of number of white pixel rows.

Horizontal Scanning is employed

# Word Separation



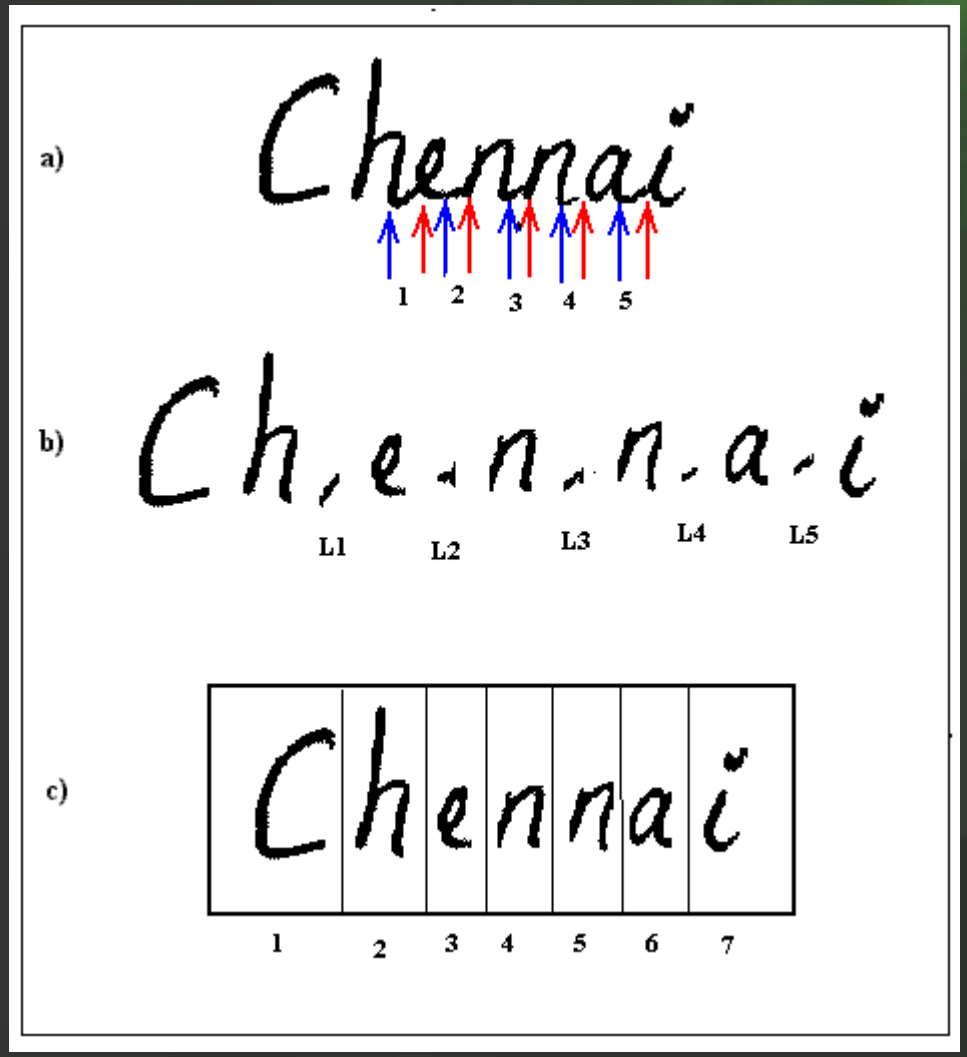

Vertical Scanning is employed

# Character Separation

a) Identification of low slopes

b) Ligature separation and removal

c) Character separation and extraction



# Address Parsing & Recognition

a) Contour formation

b) Outer Contour extraction

c) Chain-code detection

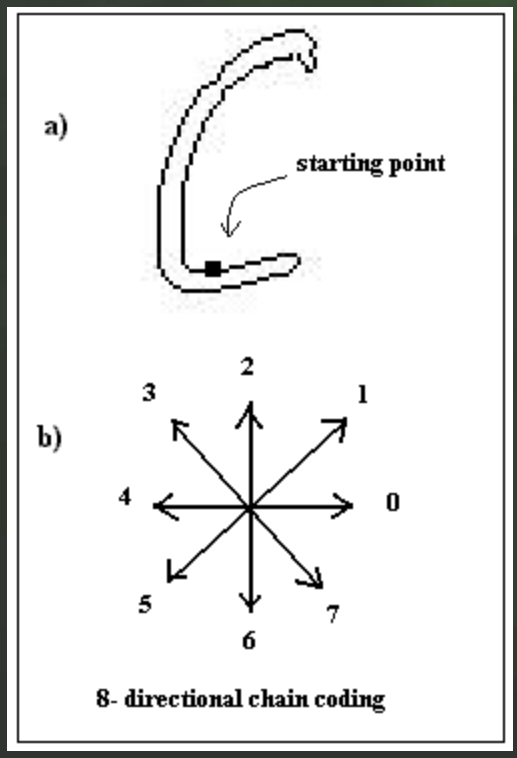
d) DFT on chain-code coordinates

$$a(u) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) \exp[-j2\pi uk/N]$$

e) Zero out the redundant freq. components

f) Match with pre-determined coeff.

g) Identify the character

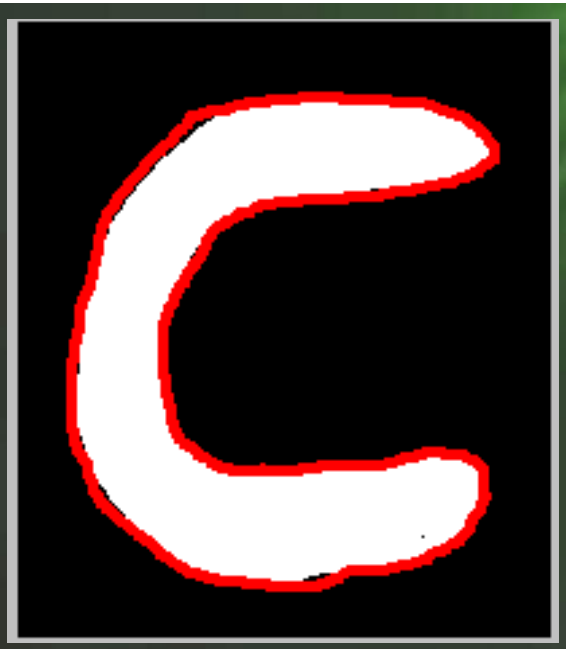


# Contour Formation

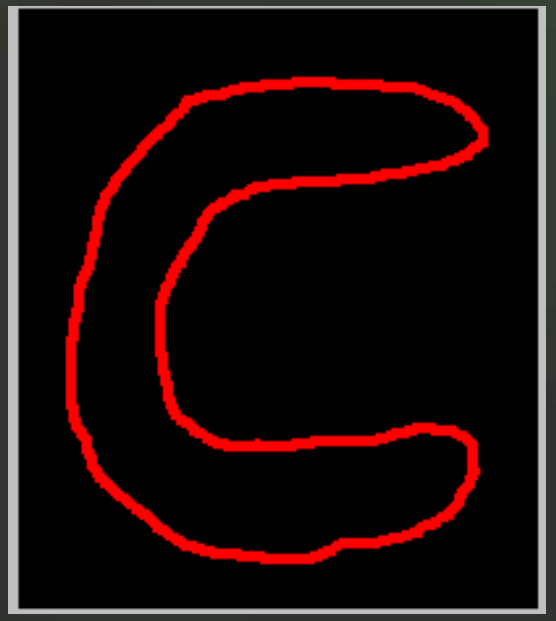
1. Connectivity
2. Horizontal Scan
3. Contour Pixels

   Sum of 4 connected pixels is less than 4

4. Contour Formation



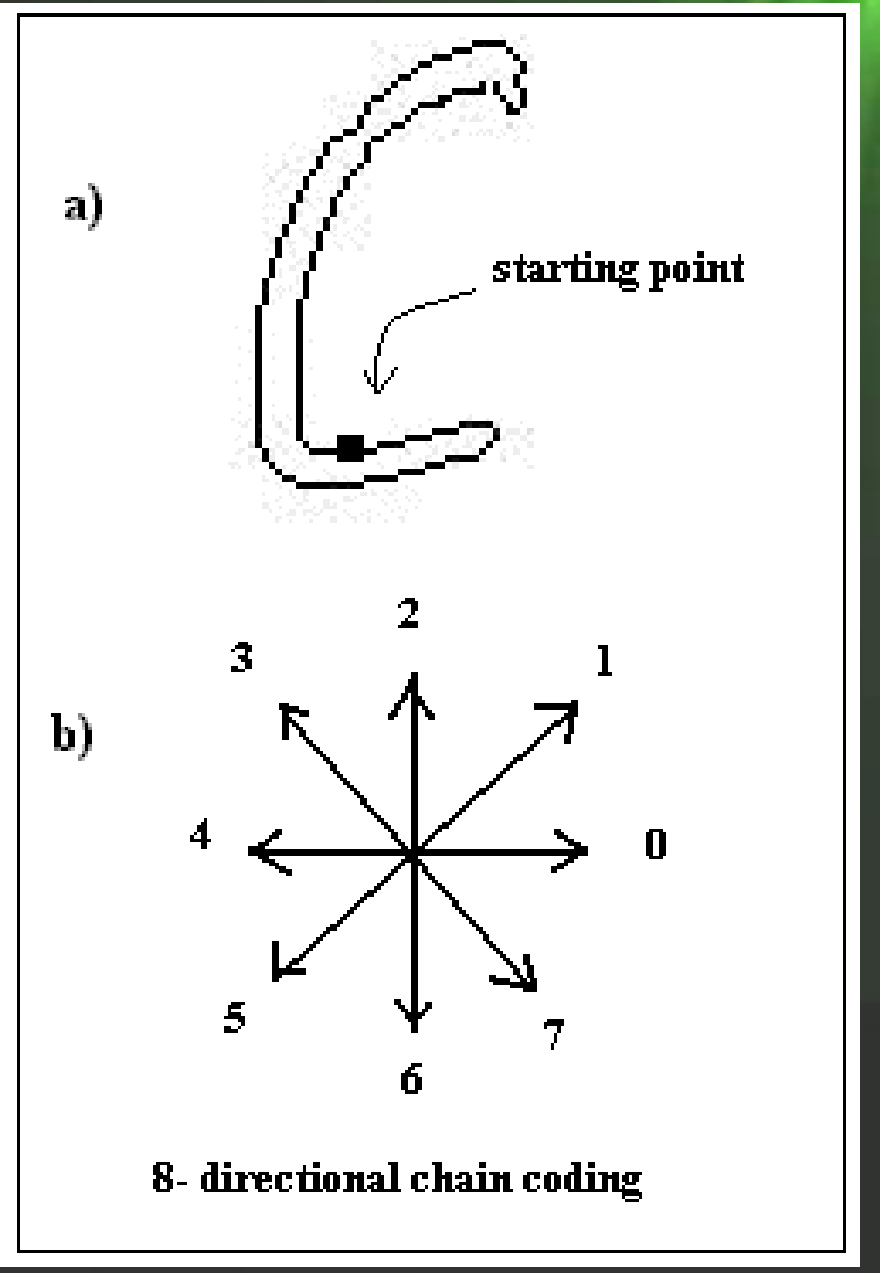
Character



Contour

# Chaincoding

1. What is a chaincode ?
2. Starting Point Location.
3. Anti-clockwise traversal
4. Formation of chaincode matrix.



# DFT Coefficients

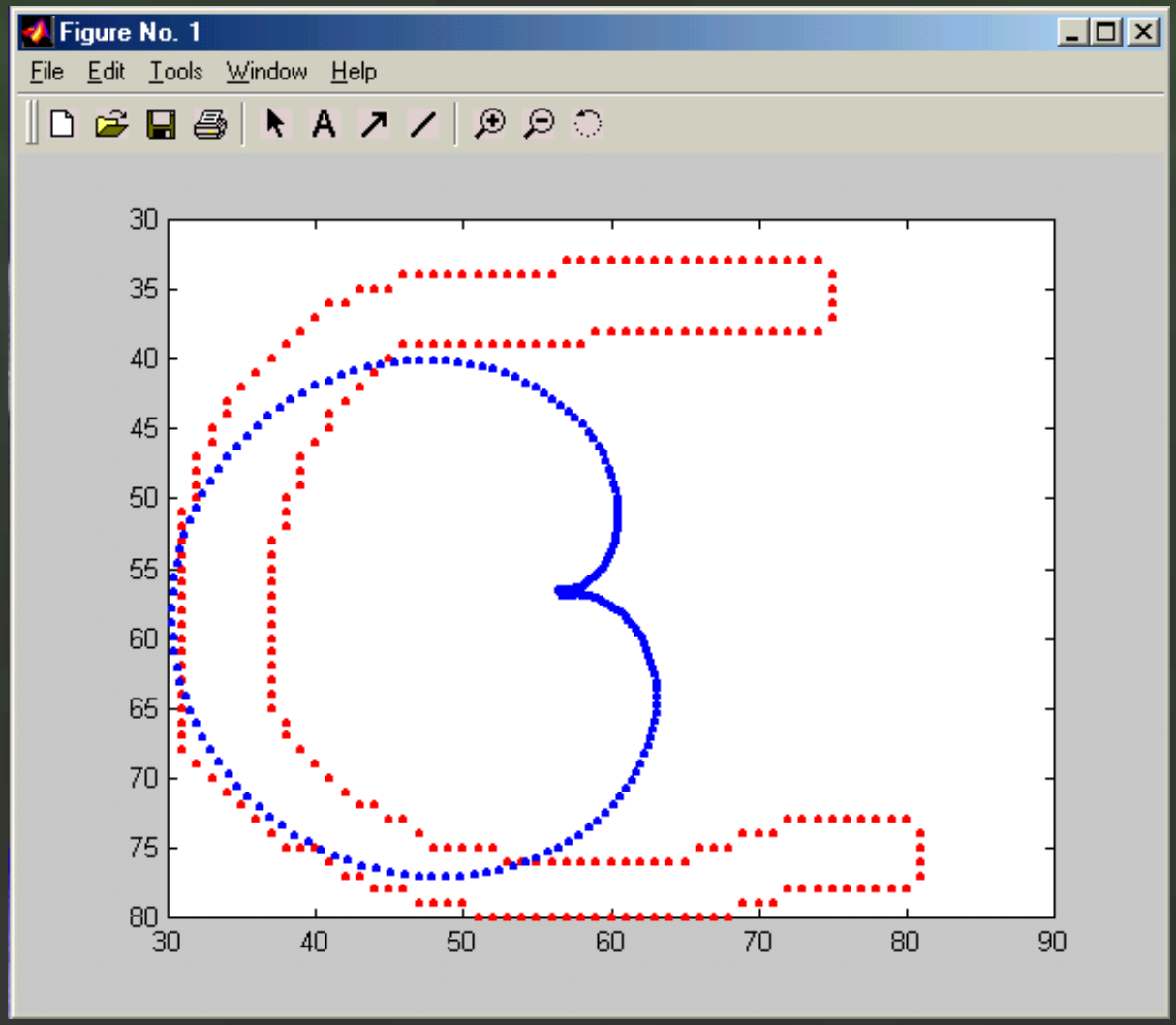
- Complex Input
- Real Part = X – Coordinate
- Imaginary Part = Y – Coordinate
- Apply DFT

$$a(u) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) \exp[-j2\pi uk/N]$$

$$S(k) = x + yi$$

➢ Choose 10 high frequency and 10 low frequency components.

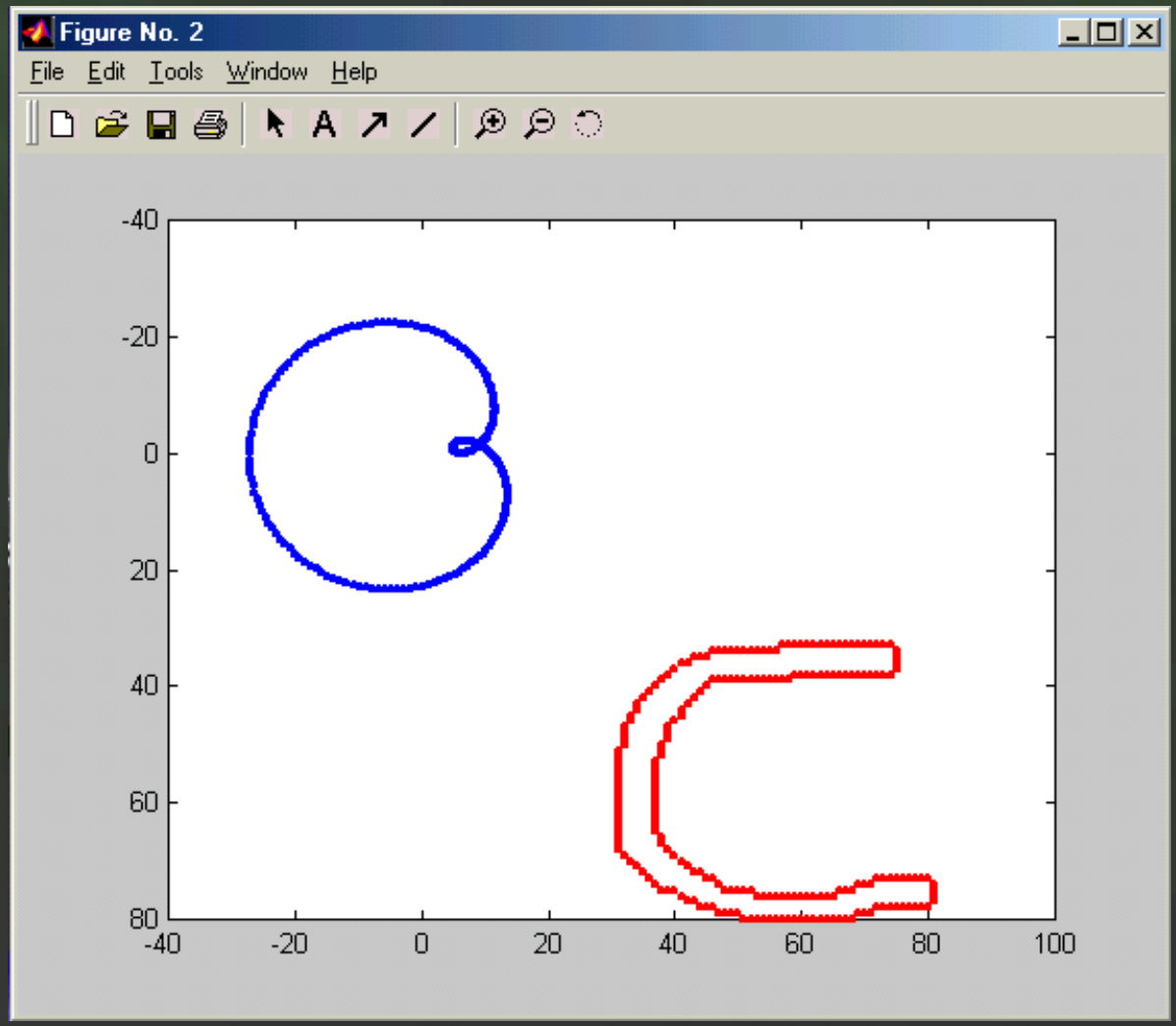
# Output with low freq components only




RED coloured ‘C’ represents the contour of original character image and

BLUE coloured ‘C’ represents the reconstructed contour with 10 low freq complex componets of DFT

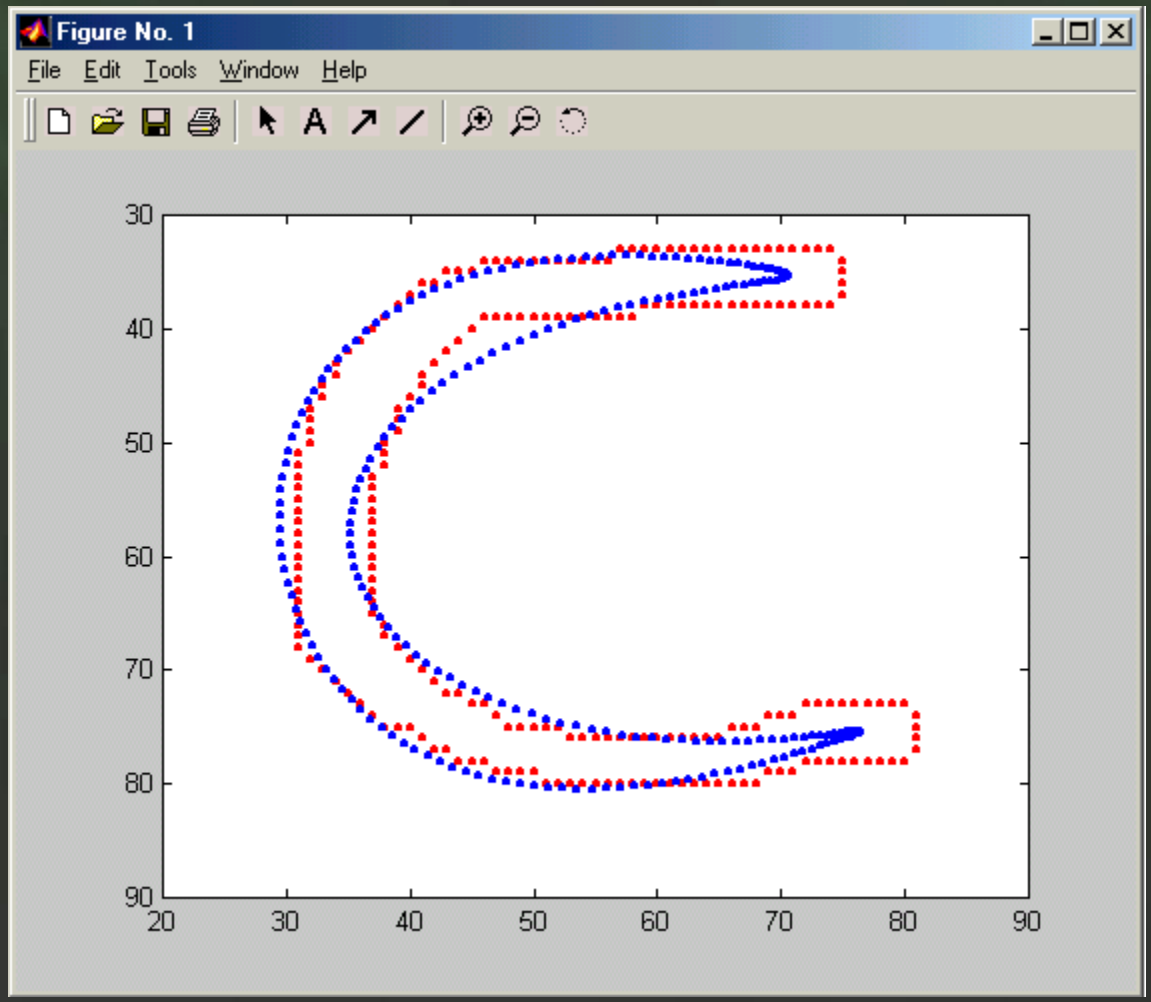
# Output with high freq components only




RED coloured ‘C’ represents the contour of original character image and

BLUE coloured ‘C’ represents the reconstructed contour with 10 high freq complex componets of DFT

# Output with both low and high freq components




RED coloured ‘C’ represents the contour of original character image and

BLUE coloured ‘C’ represents the reconstructed contour with 20 complex componets of DFT

# The RECOGNITION procedure

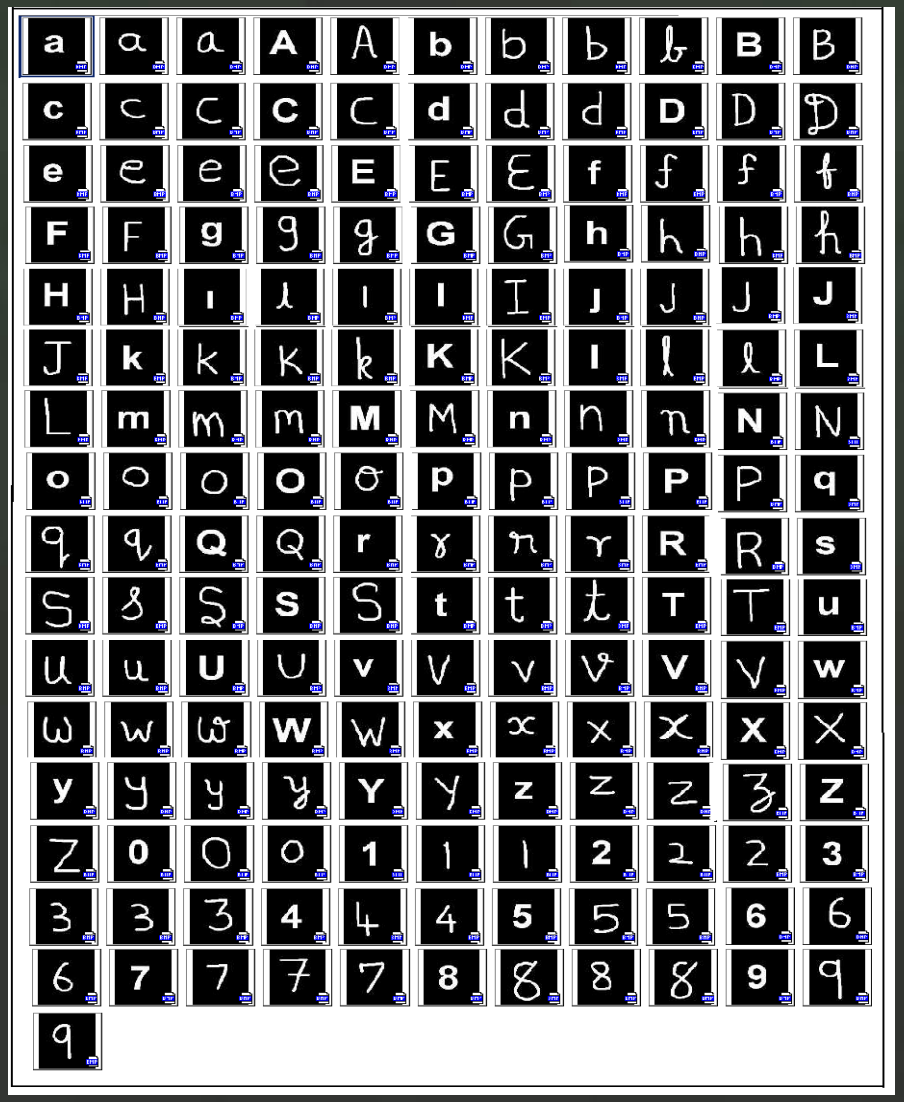
- 20 complex components for each character.
- Create complex patterns for all kinds of characters and numbers.
- This project has taken 177 kinds of character inputs.
- Use a BPN

```
a1 =

  1.0e+004 *

   1.1288 + 1.1091i
  -0.0300 - 0.0299i
  -0.0584 - 0.0025i
   0.0197 + 0.0503i
   0.0142 + 0.0145i
   0.0005 - 0.0192i
  -0.0015 - 0.0059i
  -0.0046 + 0.0063i
   0.0037 + 0.0017i
   0.0030 - 0.0040i
   0.0007 - 0.0004i
  -0.0001 + 0.0069i
   0.0044 - 0.0023i
   0.0116 - 0.0006i
   0.0100 + 0.0074i
  -0.0039 - 0.0166i
   0.0101 + 0.0031i
  -0.0233 - 0.0631i
  -0.1512 - 0.0090i
   0.0927 - 0.2816i
```

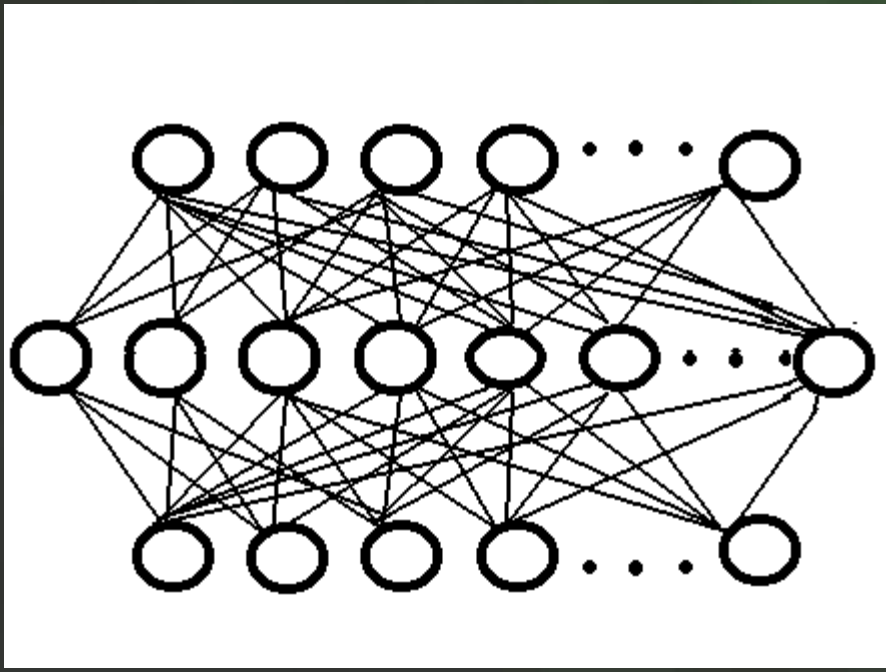
# The RECOGNITION procedure (Contd.)



177 character patterns used for training

# The RECOGNITION procedure (Contd.)

- Back Propagation Neural Network.
- Training Functions.
- TRAINGDX → Gradient Descent method.
- BPN Layers.
- Inputs to the Neural Network.
- Complexity in Complex inputs.
- Need for choosing 40 input nodes.



Neural Network

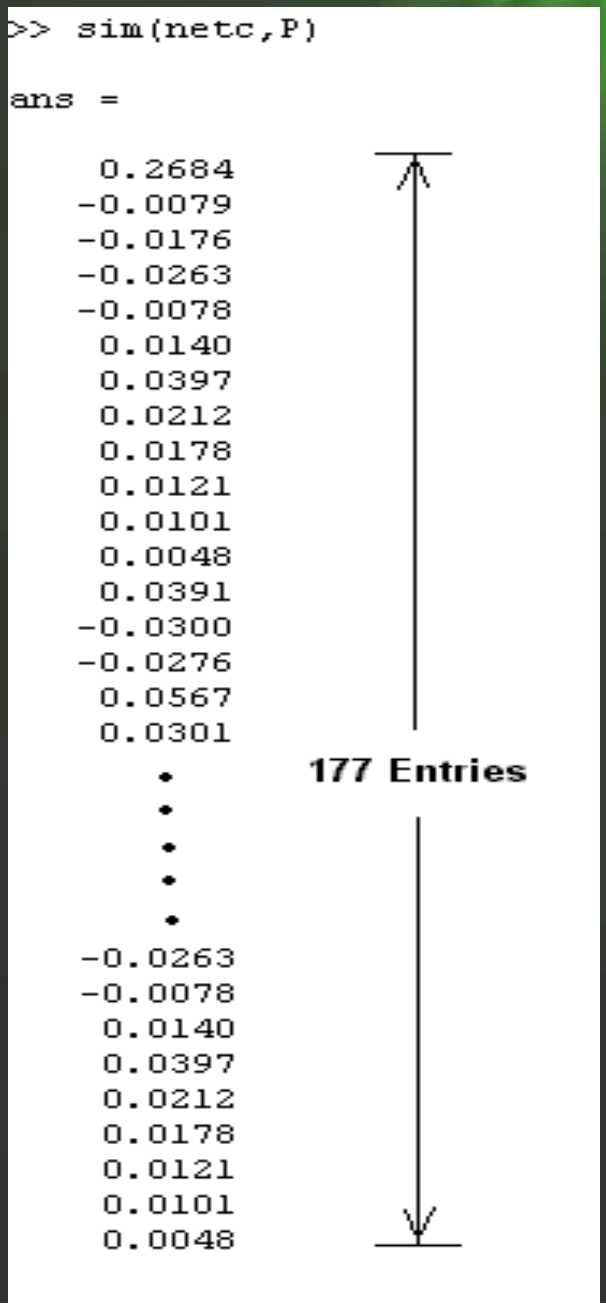
# BPN Configuration used

- No. of Layers = 3

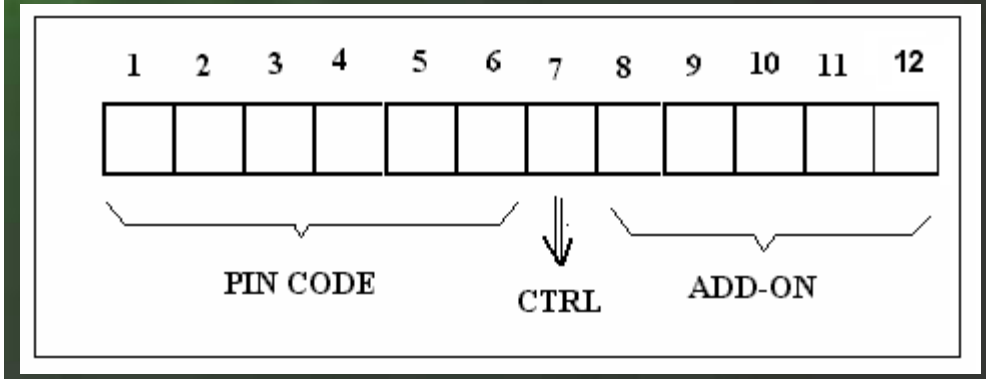
  (i.e.) No. Hidden Layers = 1
- No. of Input Nodes = 40
- No. of Hidden Nodes = 5000
- No. of Output Nodes = 177
- No. of Epochs = 3000
- Goal (MSE) = 0
- Learning Rate = 0.01
- Training function = TRAINGDX

# The RECOGNITION procedure (Contd.)

- Obtain the 20 complex DFT components from the input image fed into the Recognition module.
- Feed the Neural Network with the 40 inputs.
- The output is a vector of 177 entries of which only one output node clearly dominates.
- Return the Character / Number corresponding to the output node.



# Delivery Point Code (DPC) Generation




→ 12 Digit number

→ PINCODE + Control Digit + Add-On
(6) (1) (5)

| CTRL | Function |
|---|---|
| 0 | Add-On represents Street code |
| 1 | Add-On represents PO BOX number |
| 2 | Error: Unable to determine Add-On |
| 3 to 9 | Expansion Slots |

CTRL Combinations

S. Balaji
PO Box No: 27711
Salem - 636001

An address format

# Delivery Point Code Generation (Contd.)

- Every HPO will have a database of streets pertaining to its region.
- Street Code is obtained from this database.
- DPC is formed by appending the CTRL digit and Street Code to the PIN Code.

```
dpc =

641013000001
```

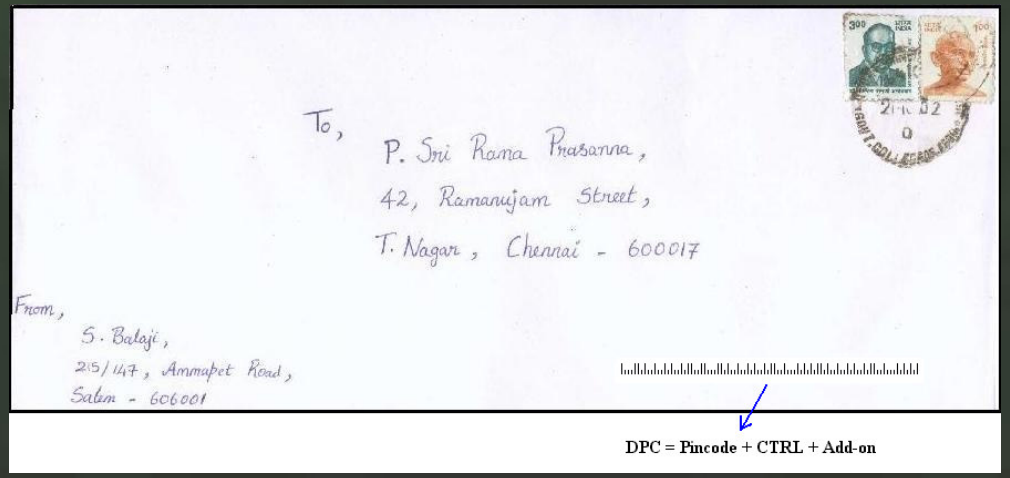
# Barcode Generation

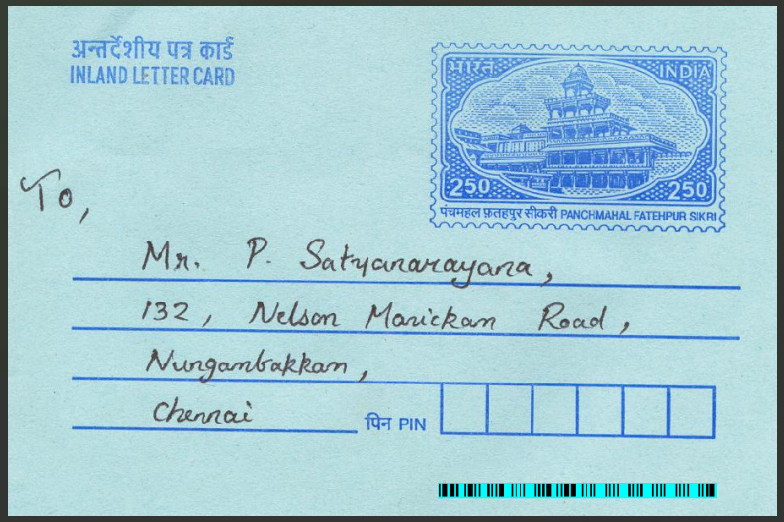
- The barcodes for digits are shown below.

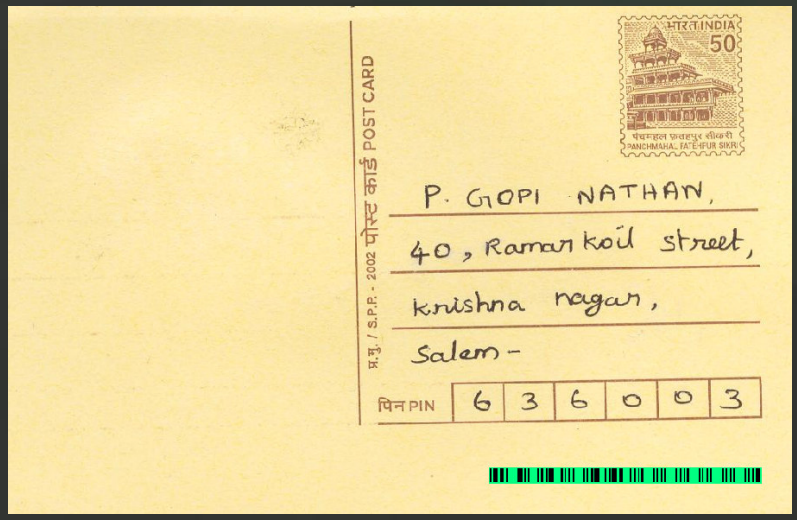

- An example Barcode for DPC = 641013000001.

# Types of DPC assignment

## a) Universal Networking method

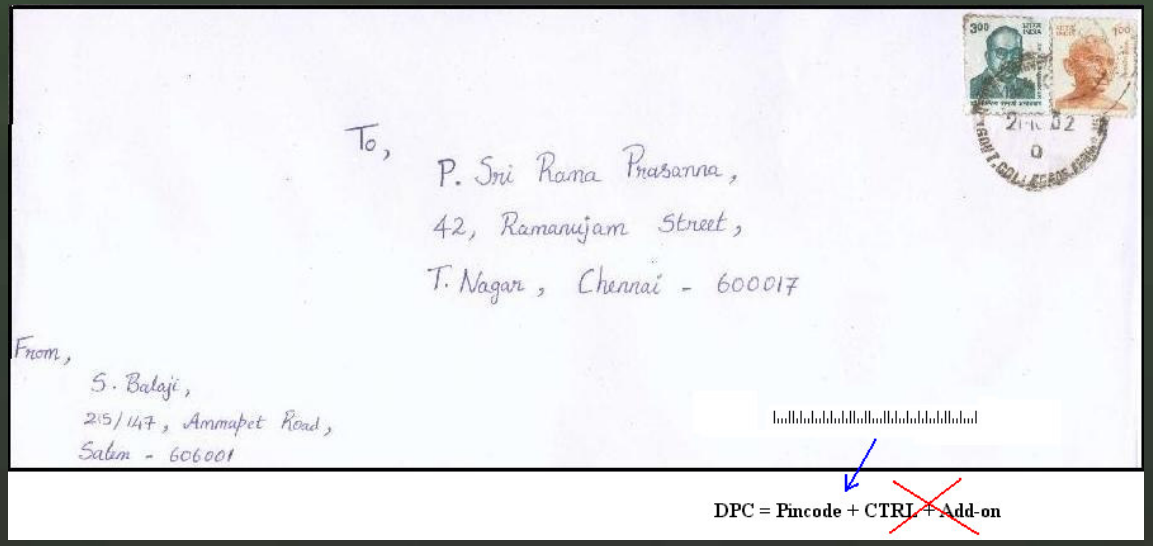


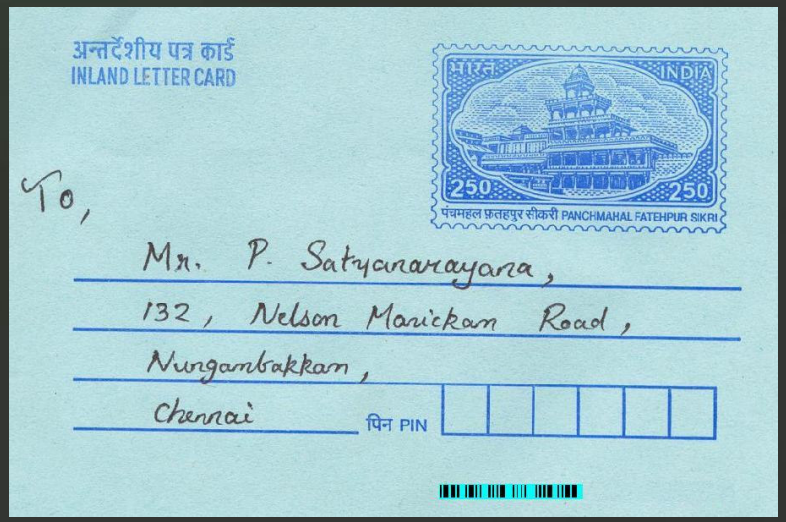



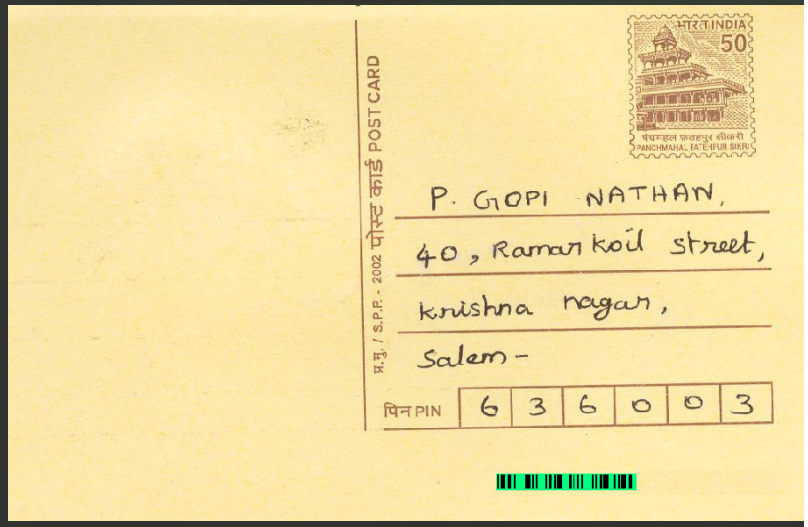



# Types of DPC assignment

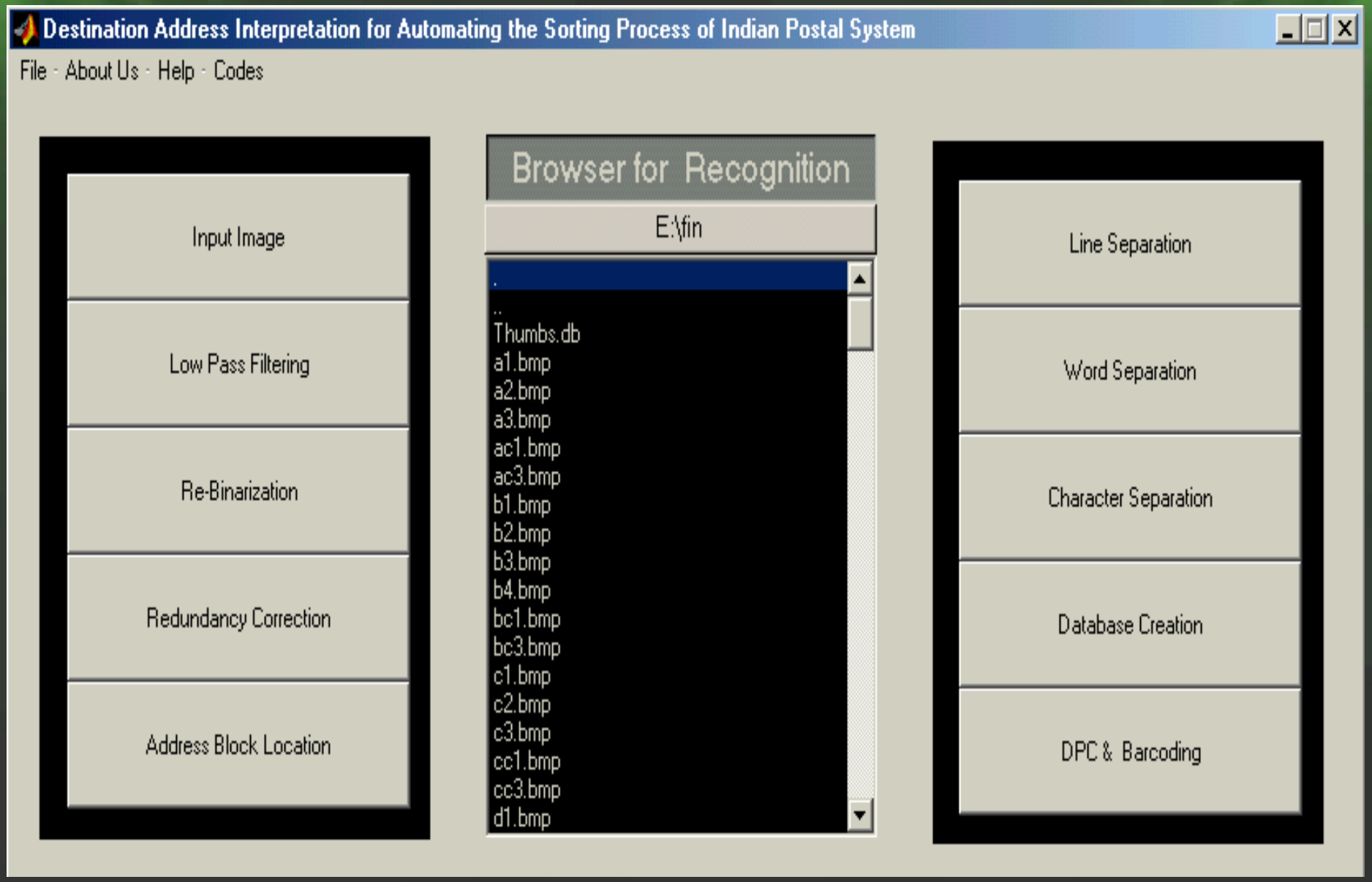
## b) Hierarchical Routing method



# Front End


Destination Address Interpretation for Automating the Sorting Process of Indian Postal System
File - About Us - Help - Codes
Input Image
Low Pass Filtering
Re-Binarization
Redundancy Correction
Address Block Location
Browser for Recognition
E:\fin
.
..
Thumbs.db
a1.bmp
a2.bmp
a3.bmp
ac1.bmp
ac3.bmp
b1.bmp
b2.bmp
b3.bmp
b4.bmp
bc1.bmp
bc3.bmp
c1.bmp
c2.bmp
c3.bmp
cc1.bmp
cc3.bmp
d1.bmp
Line Separation
Word Separation
Character Separation
Database Creation
DPC & Barcoding

# Accolades

- ✓ FIRST Prize in Computer Society Of India's Students conference at Regional Engineering College, Surathkal. - April 2003
- ✓ FIRST prize in Project presentation held at Thiagarajar College of Engineering, Madurai - March 2003
- ✓ FIRST prize in Project presentation held at Government College of Engineering, Salem - February 2003
- ✓ FIRST prize in Project presentation held at Annamalai University Chidambaram - March 2003.
- ✓ FIRST prize in Project presentation held at Tamil Nadu College of Engineering, Coimbatore - February 2003.
- ✓ SECOND prize in Project presentation held at AK College of Engineering, Krishnan Koil - March 2003

# Conclusion

- → Prospects of this project
- → Innovative Features

# References


- Interpretation of Handwritten addresses in US Mail Stream', Sargur N. Srihari, Venu Govindarajulu, Ajay Shekhawat, (CEDAR), NY 14228.
- Rafael C. Gonzalez, Richard E.Woods, 'Digital Image Processing',Pearson Edication Asia, 1992 edition.
- Anil K. Jain,'Fundamentals of Digital Image Processing', EEE,1989 edition.
- Dwayne Phillips, "Image Processing in C", BPB Publications, 1995
- Nick Efford, "Digital Image Processing", Pearson Education,2000

# Implementation in MATLAB R12


Thank You