# Table of Contents

<table>
<thead>
<tr>
<th>S.No</th>
<th>Description</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><strong>Introduction</strong></td>
<td>001</td>
</tr>
<tr>
<td>1.1</td>
<td>Preamble</td>
<td>001</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Target Research Work</td>
<td>003</td>
</tr>
<tr>
<td>1.2</td>
<td><strong>Image Processing</strong></td>
<td>004</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Digital Image definition</td>
<td>005</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Document Images Processing</td>
<td>007</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Digitization of the Data</td>
<td>013</td>
</tr>
<tr>
<td>1.2.4</td>
<td>Low-Level Processing</td>
<td>014</td>
</tr>
<tr>
<td>1.2.5</td>
<td>Segmentation</td>
<td>019</td>
</tr>
<tr>
<td>1.2.6</td>
<td>Skew Estimation and Correction</td>
<td>023</td>
</tr>
<tr>
<td>1.2.7</td>
<td>Optical Character Recognition</td>
<td>027</td>
</tr>
<tr>
<td>1.2.8</td>
<td>Graphical Documents</td>
<td>028</td>
</tr>
<tr>
<td>1.2.9</td>
<td>Data Security</td>
<td>030</td>
</tr>
<tr>
<td>1.2.10</td>
<td><strong>Information System: Definition</strong></td>
<td>033</td>
</tr>
<tr>
<td>1.3</td>
<td><strong>Statement of the Problem</strong></td>
<td>034</td>
</tr>
<tr>
<td>1.4</td>
<td><strong>Organization of the Thesis</strong></td>
<td>036</td>
</tr>
<tr>
<td>1.5</td>
<td>Contributions Made</td>
<td>038</td>
</tr>
<tr>
<td>2.</td>
<td><strong>Supporting Research Literature</strong></td>
<td>039</td>
</tr>
<tr>
<td>2.1</td>
<td><strong>Background</strong></td>
<td>039</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Block classification approaches</td>
<td>040</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Text Segmentation Approaches</td>
<td>040</td>
</tr>
</tbody>
</table>
2.1.3 Segmentation/Classification mixed

2.1.4 Page Segmentation Approaches

2.2 Block Classification Approaches

2.2.1 Linear Discriminant classifiers

2.2.2 Binary Classification Tree

2.2.3 Neural Network Classifiers

2.3 Text Segmentation Approaches

2.3.1 Connected Components Analysis

2.3.2 Projection Profile Methods

2.3.3 Texture Based or Local Analysis

2.3.4 Analysis of the Background Structure

2.4 Segmentation/Classification Mixed Approach

2.4.1 Connected Components Analysis

2.4.2 Smearing Based Techniques

2.4.3 Texture Based/Local Analysis Methods

2.5 Page Segmentation Approaches

2.5.1 Smearing Based Techniques

2.5.2 Projection Profile Methods

2.5.3 Texture Based or Local Analysis

2.5.4 Analysis of the Background Structure

2.6 Data Security

2.7 Data Integration

3. Cognition for the Recognition of Documents

3.1 Background

3.2 Pre-processing

3.3 Feature Extraction

3.3.1 Image Representation
3.3.2 Segmentation  
3.3.3 Simple model  
3.3.4 Processing Tables  
3.3.5 Text /Graphics Algorithm

3.4 Form Learning  
3.5 Document Recognition  
3.6 Conclusion

4. Intelligent Recognition of Document Types

4.1 Background  
4.2 The Proposition  
4.3 Framework

  4.3.1 The Sequence of Steps  
  4.3.2 The Domain Specific Classification  
  4.3.3 The Document Space  
  4.3.4 The set of document classes

4.4 Preprocessing and Appearance Based Decisions

  4.4.1 Preprocessing  
  4.4.2 Image representation  
  4.4.3 Appearance Based Feature Extraction

4.5 Segmentation and Feature Extraction

  4.5.1 Segmentation  
  4.5.2 Preserving the Spatial Relationships amongst Segments

4.6 Building of a Knowledge Base and Assignment of Class Labels

  4.6.1 Block Distances
4.7 Assignment of Class Labels and Document Recognition

4.8 Conclusion

5. Extraction of Dynamic Variants

5.1 Background

5.2 System Overview

5.3 Issues in extraction of Data
   5.3.1 Many form formats
   5.3.2 Errors in typing information onto a form
   5.3.3 Poor typing quality
   5.3.4 Many fonts
   5.3.5 Scanner mis-registration
   5.3.6 Extraction Procedure
   5.4.7 Extraction Algorithm

5.5 Table Processing

5.6 Data Storage in the Database

5.7 Conclusion

6. Secured Environment for Data Extraction

6.1 Background

6.2 Access to Database Systems

6.3 Encryption & Decryption

6.4 How Encryption works
   6.4.1 Encryption Key Management
   6.4.2 Encryption Types

6.5 Protection with Encryption

6.6 Risk Calculation
6.7 Proposed System

6.8 Implementation

6.9 Interface and tools for the Application Package
  6.9.1 Encryption/Decryption Layer
  6.9.2 Interface and tools for Database

6.10 Case Study

6.11 Security of the Extracted Data

6.12 Where & When to insert the Decrypt Key

6.13 Conclusion

7. System Integration for Efficient Data Handling
  7.1 Background
  7.2 Methodology
  7.3 Database Analysis
  7.4 Current Database Configuration
  7.5 Database Integration Design
  7.6 Data Warehouse
  7.7 Conclusion

8. Conclusion & Future Work
  List of Publications
  Bibliography
  Appendix