

Handwritten Character Recognition using Machine Learning Approach

Mr. Shivang R. Patel
(130320702512)

Ms. Jasmine Jha (Guide)
Assistant Professor,
PGCE Department, LJJET

L.J. Institute of Engineering & Technology, Ahmedabad
Gujarat Technological University, Ahmedabad

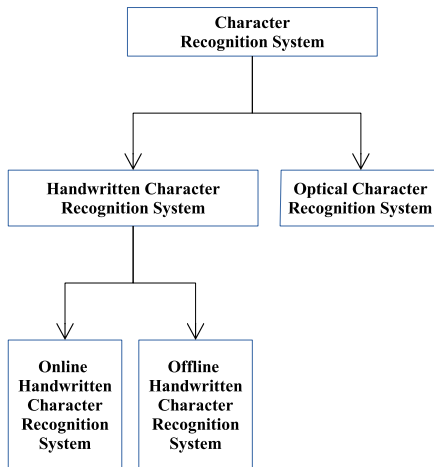


Jun 23, 2015

Outline

- 1 Introduction
- 2 Literature Review - A Comparison Table
- 3 Proposed Hypothesis
- 4 Work Done so Far...
- 5 Result & Analysis
- 6 Conclusion
- 7 References

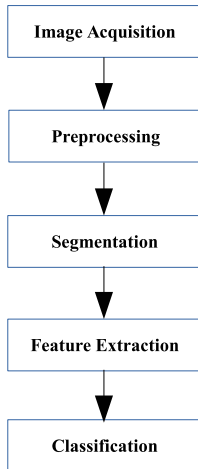
Character Recognition System



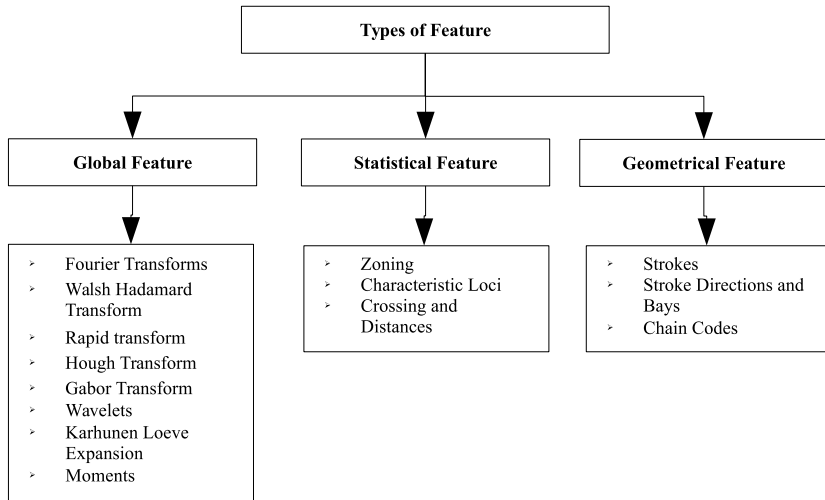
Applications

- Character identification
- Digitization of handwritten record
- Application form reading and based on data entry
- Translation system - recognize the unknown language and translate it in a known language^[1]
- Bank cheques processing^{[2], [3]}
- Signature verification
- Vehicle number plates^{[2], [3]}
- Automatic pin code reading to postal mail ^{[2], [3]}

Framework of HCR System

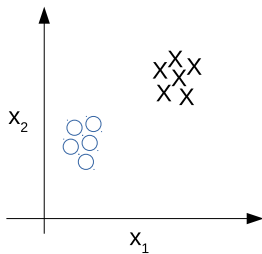


Types of Features^[4]

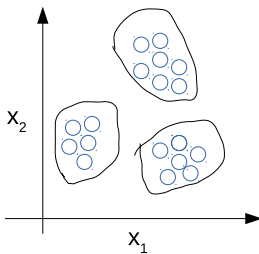


Types of Learning

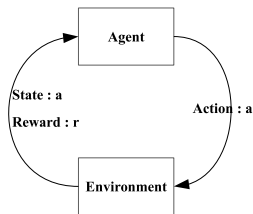
Supervised Learning



Unsupervised Learning

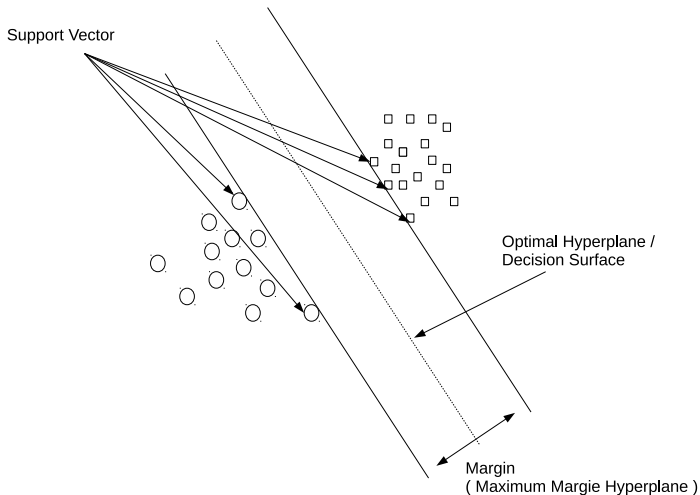


Reinforcement Learning



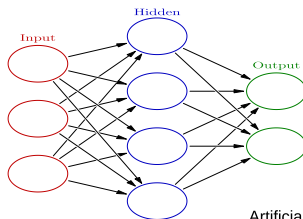
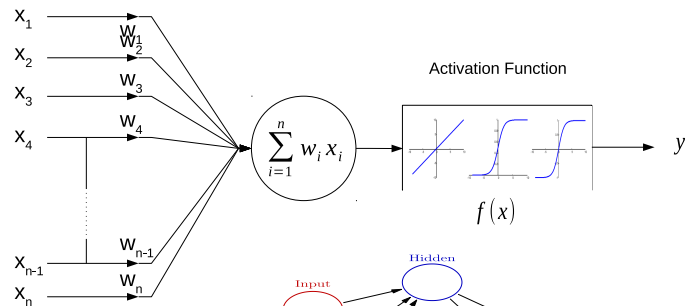
- **Lazy Learning** : Classification using Nearest Neighbors
- **Probabilistic Learning** : Classification using Naive Bayes
- **Divide and Conquer** : Classification using Decision Tree and Rule
- **Black Box Methods** : *SVM and Artificial NN*
- **Finding Pattern** : Market Basket Analysis using Association Rules

Support Vector Machine - SVM



Single Neuron - ANN

The McCulloch-Pitts Model of Neuron



Artificial Neural Network

Literature Review - A Comparison Table

Paper	Features Extraction	Language	Dataset	Classifier	Discussion
1 ^[5]	Grapheme segmentation and Sliding Window	English	Rimes Database	MLP	Very Fast But Low accuracy
2 ^[6]	DWT with Multi resolution technique	English Characters	Own Character Dataset	NN with Euclidean Distance matrix	good accuracy – up to 99.23% more time
3 ^[7]	Multi zoning, Geometrical feature distance and angle, topological feature end point transition, Directional feature chain code histogram	English characters and symbols	Own dataset	BPNN and SVM	BPNN: 98% for numeral, 96.5% for special characters, 95.35% for uppercase characters and 92% for lowercase characters SVM: 92.167% for both case characters

Human Learning...

Learning I : We have 100 things to learn at a time !

Learning II : We have 10 things to learn at a time !

Result : *Obviously, Learning II have better accuracy and less time then Learning I !*

Learning III : If we divide 100 things in to 10 parts and learn it !

Result : ?!!

Problem Statement

Characters are changed with language and actual shape of character change with person to person - This things make character recognition more difficult.

Problem Statement :-

” Accuracy of Character Recognition need to be Improve with lower Computation (time reduction for recognition).”

Goal of Proposed System

Goal

To develop the handwritten English characters recognition system that **Reduce** the **Computation** and improve the **Accuracy** with the use of **SVM** and **ANN** machine learning technique.

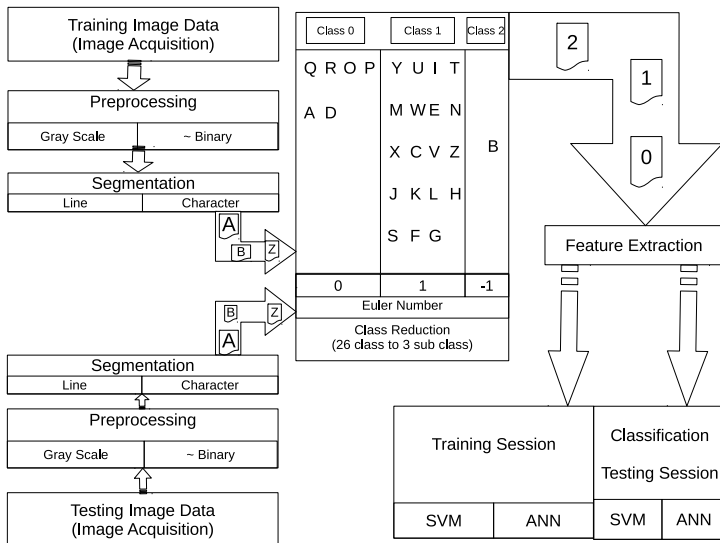
Assumptions and Constraints

- Zero noise present in available dataset. Data is cleaned.
- Not supported for handwritten cursive writing.
- For experiment - only capital English characters.
- For the 26 character classes to subclass creation/reduction any kind of other techniques(image processing) can be used. Like, Euler number, horizontal or vertical symmetry, number of horizontal, vertical, cross, angular lines are available in characters, etc. Here, in proposed system Euler number based class reduction proposed.

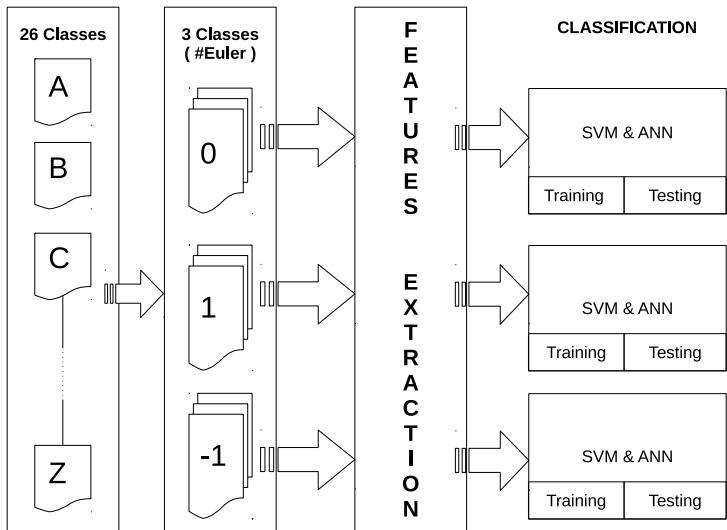
Assumptions and Constraints

- Developed system only deal with 0,1 and -1 Euler number based class. That mean, character must contain 0(one hole),1(no hole) and -1(two hole) Euler number.

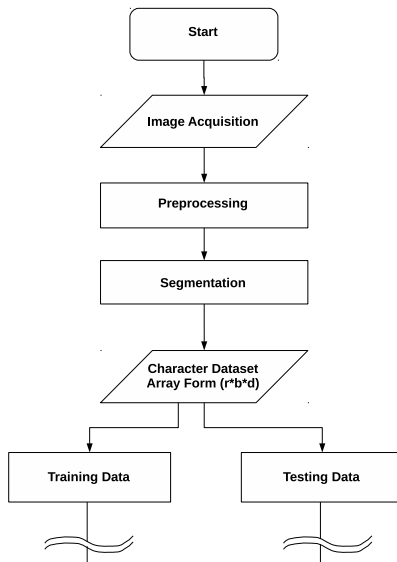
Proposed Methodology



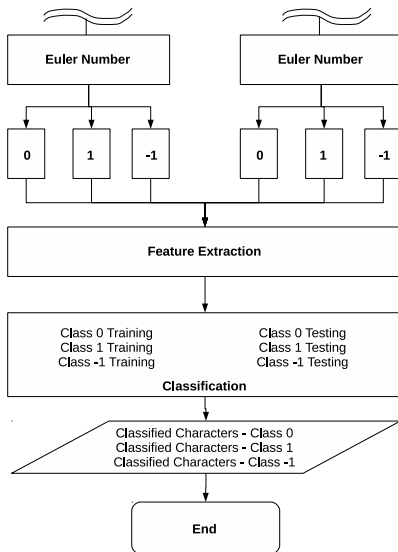
Computation Reduction



Flowchart for Proposed HCR system



Flowchart for Proposed HCR system



Tools

- Matlab (R2014a)

Technology


- Matlab

System Development Strategy

Module : 0	Prepare Experimental Dataset
Module : 1	Implement FE technique and #Euler base class reduction
Module : 2	Implement SVM – Training & Testing phase
Module : 3	Implement ANN – Training & Testing phase
Module : 4	Combine Module 0-1-2(SVM) & 0-1-3(ANN)
Work Done !	Experiment - Result & Analysis

Dataset


Dataset Detail (Total Characters 11548)





Character	Sample	Character	Sample	Character	Sample	Character	Sample	Character	Sample
A	444	G	389	M	453	S	447	Y	450
B	432	H	403	N	437	T	469	Z	468
C	518	I	818	O	459	U	457		
D	385	J	426	P	467	V	482		
E	367	K	374	Q	452	W	478		
F	70	L	497	R	434	X	472		

Feature Extraction

Table : Feature Vector - Detail

No.	Feature Extraction Technique	Fix Size / variable size	Size of Feature Vector
1	Sobel	Fix Size	60
3	Histograms of Oriented Gradients	Fix Size	288
4	Zoning	Fix Size	143
5	Chain Code	Fix size	8

Euler-number Base Class Reduction

Current Folder

- Name
- SymmetryTesting
- Testing
- Training
- automaticParameterSelection.m
- binary.m
- cell2csv.m
- classificationPhase1.m
- data.txt
- debug.mat
- demo.m

binary.m (Function)

Workspace

Name	Value	Min
class0	40x30x66 logical	
class1	40x30x209 logical	
class2	40x30x11 logical	
cntC0	67	67
cntC1	210	210
cntC2	12	12
count	286	286
d	286	286
data	40x30x286 logical	
eul	1	1

Editor - /home/shivangpatel/matlab/dissertation/HCCR/classificationPhase1.m

```
demo.m x classificationPhase1.m x +
24 - class1(:, :, cntC1) = cat(3, data(:, :, count));
25 - cntC1 = cntC1 + 1;
26 - end
27
28 %% % class 0 ---- 1 holes...euler 0
29 - if (eul == 0)
30 -     class0(:, :, cntC0) = cat(3, data(:, :, count));
31 -     cntC0 = cntC0 + 1;
32 - end
33
34 File Tools Window Help
35
36
37
38
39
40
41
42 Pixel info: (X, Y) BW
```

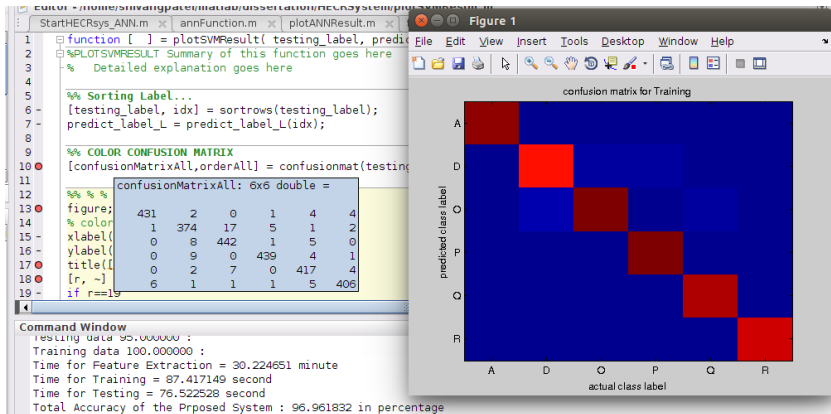
Image Tool 1 - (MATLAB Expression)

100%

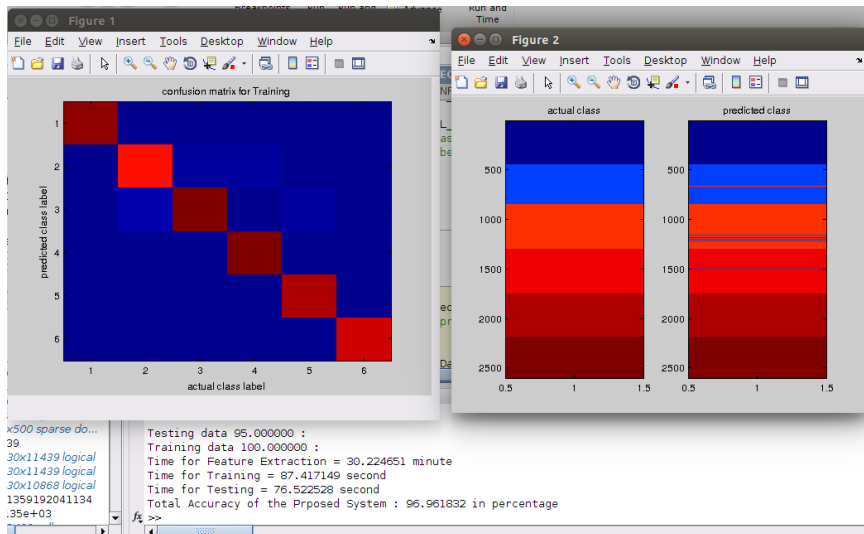
Command Window

```
fprintf('Total #char in dataset is :%d \n', d);
fprintf('Total Elements in Class 0 is :%d \n', cntC0-1);
fprintf('Total Elements in Class 1 is :%d \n', cntC1-1);
fprintf('Total Elements in Class 2 is :%d \n', cntC2-1);
Total #char in dataset is :286
Total Elements in Class 0 is :66
Total Elements in Class 1 is :209
Total Elements in Class 2 is :11
>> imtool(class2(:, :, []))
>>
```

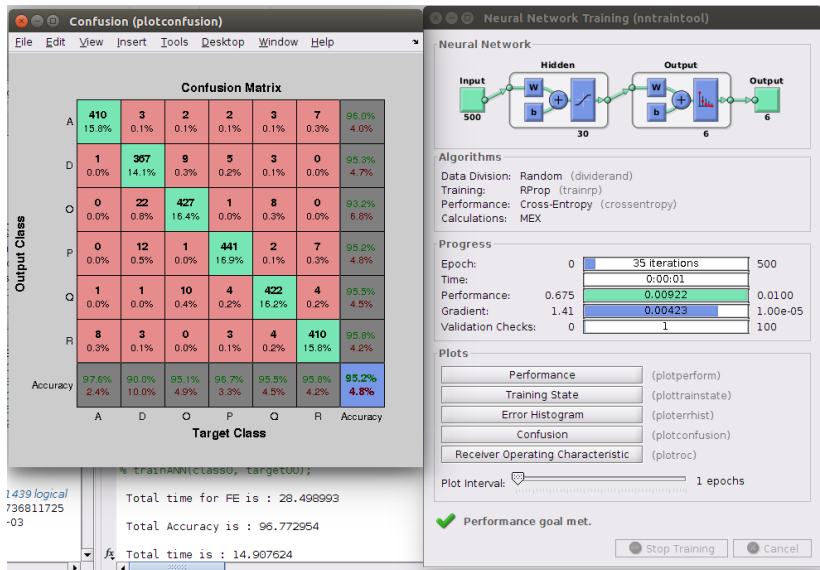
SVM Training and Testing



SVM Training and Testing



ANN Training and Testing



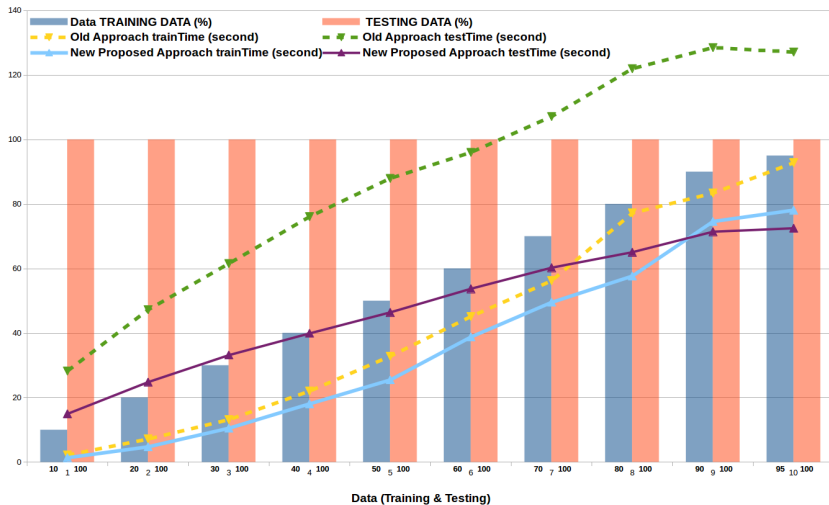
Existing approach vs Proposed approach - SVM

(Accuracy Parameter)



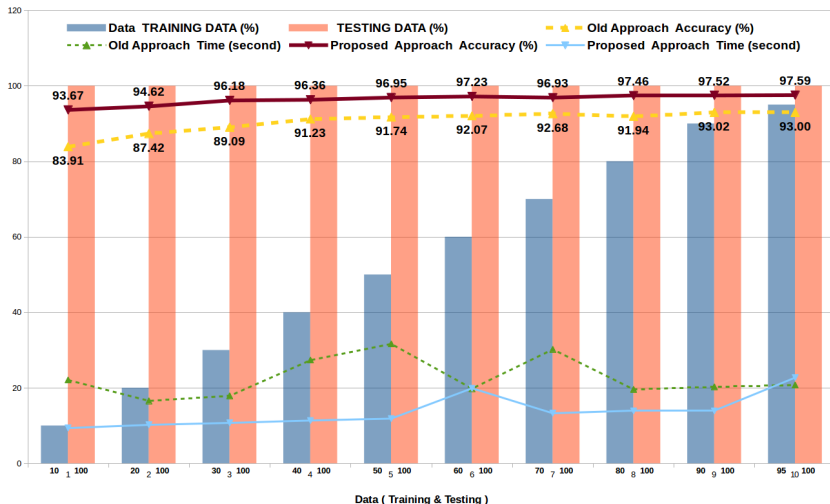
Existing approach vs Proposed approach - SVM

(Time Parameter)



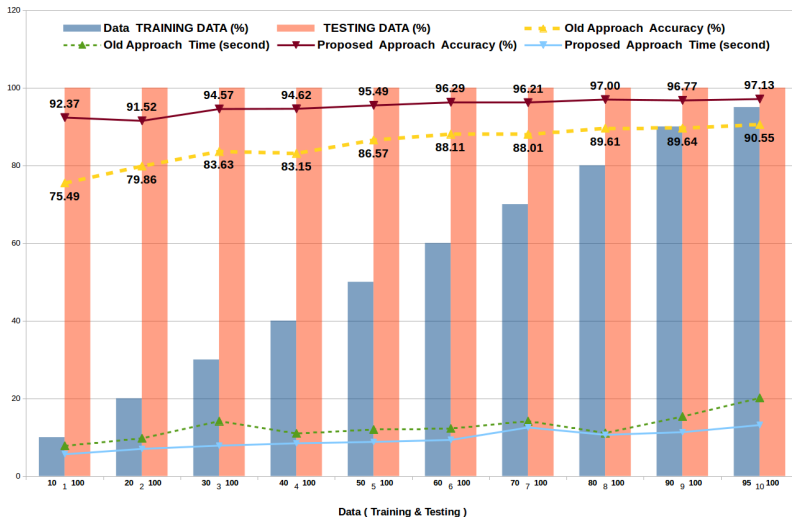
Existing approach vs Proposed approach - ANN

(Accuracy & Time Parameter - "trainscg")



Existing approach vs Proposed approach - ANN

(Accuracy & Time Parameter - "trainrp")



Conclusion

Applying machine learning techniques in handwritten character recognition system increase the recognition accuracy and selection of good features reduce the time complexity of HCR system. All result show that proposed system quite good then conventional system approach. It take less time and gives overall good accuracy.

Future works : In future work, remove as possible as constraints and experiment again this novel approach with different handwritten dataset.

References

- [1] SA Angadi and MM Kodabagi. “A Robust Segmentation Technique for Line, Word and Character Extraction from Kannada Text in Low Resolution Display Board Images”. In: *Signal and Image Processing (ICSIP), 2014 Fifth International Conference on*. IEEE. 2014, pp. 42–49.
- [2] SV Rajashekararadhya and P Vanaja Ranjan. “Zone-based hybrid feature extraction algorithm for handwritten numeral recognition of two popular Indian scripts”. In: *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*. IEEE. 2009, pp. 526–530.
- [3] Laurence Likforman-Sulem. “Recent Approaches in Handwriting Recognition with Markovian Modelling and Recurrent Neural Networks”. In: *Recent Advances of Neural Network Models and Applications*. Springer, 2014, pp. 261–267.

References

- [4] Gaurav Kumar and Pradeep Kumar Bhatia. “A Detailed Review of Feature Extraction in Image Processing Systems”. In: *Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on*. IEEE. 2014, pp. 5–12.
- [5] Theodore Bluche, Hermann Ney, and Christopher Kermorvant. “Feature extraction with convolutional neural networks for handwritten word recognition”. In: *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. IEEE. 2013, pp. 285–289.
- [6] DK Patel, T Som, and Manoj Kumar Singh. “Multiresolution technique to handwritten English character recognition using learning rule and Euclidean distance metric”. In: *Signal Processing and Communication (ICSC), 2013 International Conference on*. IEEE. 2013, pp. 207–212.

References

- [7] Nisha Sharma, Bhupendra Kumar, and Vandita Singh. “Recognition of off-line hand printed English Characters, Numerals and Special Symbols”. In: *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-*. IEEE. 2014, pp. 640–645.

Thank You !

