International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 4
ISSN: 2454-4248
599 – 602

Character Recognition System using Radial Features

Abhisehek Anand and Pankaj Bhambri
Department of Information and Technology,
Guru Nanak Dev Engineering College, Gill Road, Ludhiana (Punjab), India

Abstract-- Extraction of text from documented images finds application in maximum entries which are document related in offices. The most of the popular applications which we find in public or college libraries where the entries of number of books are done by manually typing the title of book along with other credentials like name of the author and other attributes. The complete process can be made effortless with the application of a suitable algorithm or application software which can be extract the documented part from the cover of book and other parts of the book thereby reducing the manual job like typing of user. Which reduces the overall job to only arranging the book title etc.by formatting the material.

*****

I. INTRODUCTION

Radial features includes the radial profile of a pattern extracted around the centre of mass of the pattern under test. The centre of mass is extracted using the second order moments of Cartesian coordinates and radii in all four quadrants are used to describe and recognize the given pattern. The same features may be used to train a classifier in order to classify the given pattern in different classes.

II. LITRATURE SURVEY

Giacinto G. et. al. (2007) This paper narrates the experimental results of the application of various pattern recognition algorithms which leads to the advancement of earthquake risk for the geological structures. [1] Aki Vehtari et. al. (2000) This paper exhibit the benefits of exercising image analysis by Bayesian multi-layer perceptron (MLP) neural networks. The Bayesian approach delivers constant way to do implication by merging the confirmation from the data to erstwhile information from the problem. [2] Enrique Frias-Martinez et. al.(2001) The problem of Signature verifications is more recognized as compare to the automatic signature recognition despite the fact that automatic signature recognition is a potential application for processing historical and legal documents and accessing security-sensitive facilities. [3] Kamal Nasreddine et al. (2009) In this paper we characterize a separation between shapes in light of geodesics fit as a fiddle space. The proposed separation, hearty to anomalies, utilizes shape coordinating to analyse shapes privately. Multi scale examination is acquainted all together with evade issues of neighbourhood and worldwide variability. [4] B. Lerner et al. (1996) Highlight extraction has been dependably commonly concentrated on for exploratory information projection and for characterization. Highlight extraction for exploratory information projection goes for information representation by a projection of a high-dimensional space onto a few dimensional space, while highlight extraction for grouping by and large requires more than a few elements.[5] Yas Abbas Alsultanny et al. (2002) The hereditary calculation actualized with neural system to decide consequently the for information representation by a projection of a high-dimensional space onto a few dimensional space, while highlight extraction for grouping by and large requires more than a few elements.[5] Yas Abbas Alsultanny et al. (2002) The hereditary calculation actualized with neural system to decide consequently the suitable system building design and the arrangement of parameters conclusion apparatus, is being enhanced to be more touchy to little changes and to incorporate more insight to handle dynamic procedure data. [8] Anil K. Jain et al. (2000). The essential objective of example acknowledgment is administered or unsupervised grouping. Among the different structures in which design acknowledgment has been customarily detailed, the measurable methodology has been most seriously concentrated on and utilized as a part of practice. All the more as of late, neural system strategies and routines imported from factual learning hypothesis have been accepting expanding consideration. [9]

III.RELATED WORK

Very firstly, all the text characters are exposed to features extraction algorithm. Each character image is used as input image. The features are extracted using the algorithm discussed in below:

Following are the statistical features which are computed from analysing the pattern in comparison to the centre of gravity for categorically organizing:

[a] Normalised Maximum Radii in all four Quadrants is represented by R1, R2, R3, and R4. See fig. (2)

[b] With respect to the object's centre of gravity X1, X2, Y1 and Y2 are Intercepts on each axis See fig. (2).

 ISSN: 2454-4248
599 – 602

[c] The Mean Radius (RM)

[d] Figure Aspect (FA) which is ratio of length to width Here FA = (X1 + X2) / (Y1 + Y2)

[e] Normalised Perimeter (Np) Here Np is the Total no. of pixels at the contour of the object / RM

[f] Normalised Standard deviation (NSD) of radii which is taken from centre of gravity of the object.

SD= √[(Ri – RM)2 /Np] NSD = SD/ RM

Where the mean radius and ith radius is RM, and Ri respectively i.e. distance between centre of gravity and ith pixel on contour of the pattern.

[g] Normalised Area (NA) of the pattern is equal to the total pixels on objects / RM2

With respect to the mean radius of the image pattern all the features are normalized. Which makes all the pattern's statistical features size independent. To classify an object the set of statistical features can be describe as figures of merit.

R1 R2 R3 and R4 are normalized maximum radii in all four quadrants and X1 X2 Y1 and Y2 represent intercepts on each axis. When all the characters are exposed to features extraction algorithm. An array is generated consisting of all features. This array is presented to a neural network system for training purpose.

Following are the steps followed in order to achieve the objectives:

- **Image Acquisition**

  The image is acquired using the digital or web camera and stored in jpeg format i.e. colored or RGB image.

- **Image Pre-processing Operations and Get Binary Image**

The jpeg image is then converted to gray scale image (8-bit color format) and exposed to thresholding operation in order to get the binary image

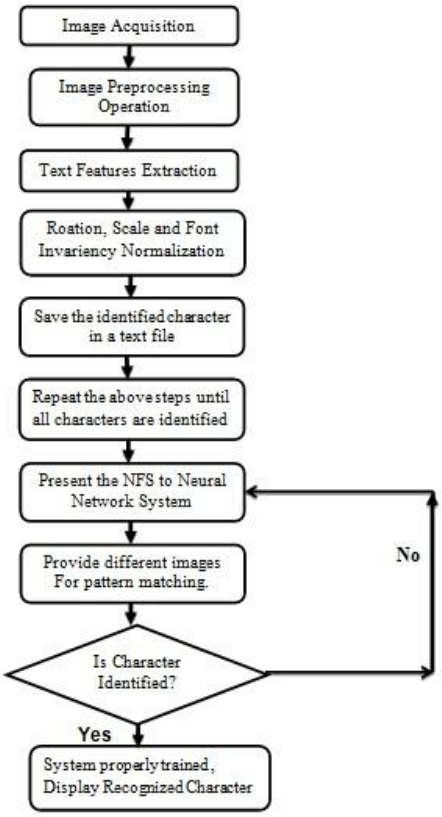

Fig 1. Flow chart of algorithm

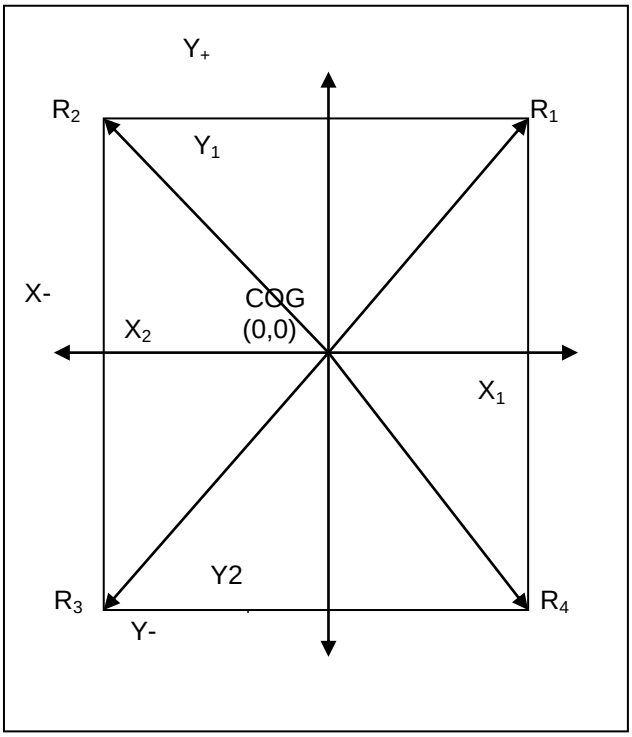

Fig 2. Centre of Gravity

 ISSN: 2454-4248
599 – 602

- **Image Segmentation**
  The binary image so obtained is now segmented for textual area identification. Once the text area is localized. The focus is shifted to text/character segmentation.

- **Textual Area Extraction**
  The textual area is extracted based on text density in a particular region.

- **Text/Characters Segmentation and Thinning**
  The text characters are segmented using the local neighbourhood connectivity of pixel suing the bwareaopen command in matlab. Further, it is important to note the distance between the characters and word. Normally, the distance between two characters of a word is less than the distance between two words. This helps in identifying the characters and text or words in the image.

- **Text Features Extraction**
  Features Normalization – Invariency Insertion

- **Location, Size and Rotation Invariency Normalization**
  The character is now made rotation independent by applying orthogonal transformation. After this, statistical features like, maximum radii in each quadrant, intercepts on each axis, perimeter, standard deviation, figure aspect and area, are computed from the image. The set of data of statistical parameters of each pattern so obtained is normalised to mean radius and stored for classification/categorization purposes.

- **Get Normalized Features Set (NFS)**
  The feature set is normalized w.r.t. size, rotation and location so that if the camera or object is varied, the features remains constant for the repeated object.

- **Present the NFS to neural Network System for Character Identification**
  The normalized feature set is presented to the neural network in order to train the system. Once the system is trained, it can be used for validation purposes. The validated system is now ready to be used in application. The identified characters are stored in a text file, from there the same may use in application.

- **Save the identified character in a text file**
  The text file is the result file that can be used for word processor.

- **Repeat the above steps until all characters are identified**
  The above steps are repeated till all characters are identified and stored in text file.

- **Present the text file as result for further processing**
  The text file so obtained is now ready to be used for word processing or we can say that is the result file.

IV. RESULTS

The result are shown for characters from A to Z and numerals 0 to 9 in table below.

V. CONCLUSION

The conferred algorithm is a noble advent in image segmentation specifically in extraction of text from the image. This can be further prolonged to algorithms used in applications in library management system, image to text converters and word processing. Some currently existing algorithm approaches to achieve the same are studied. But they came up with certain disadvantages of invariance in size, location and rotation. In the suggested algorithm, the different invariances are considered into the processes for features extraction such that the normalization of characters are done in every respect. Once the features of different characters are fixed and are consistent for the same object in both form, then that features can be practised for character recognition purposes. This algorithm is written and developed in matlab version 7.5 and can be further converted to any front end language for commercial purpose

 ISSN: 2454-4248
599 – 602

TABLE 1: RESULT FOR CHARACTERS AND NUMERALS

Character	MaxR1	MaxR2	MaxR3	MaxR4	Mmax	MinR1	MinR2	MinR3	MinR4	Mmin	MR	X1	X2	Y1	Y2	SD	Area	NP	FA	Euler
A	1.34	1.35	1.56	1.55	1.45	0.66	0.67	0.49	0.5	0.58	1	0.71	0.72	1.32	0.49	0.3	2.5	8.14	0.78	0
B	1.12	1.27	1.26	1.14	1.2	0.56	0.76	0.76	0.73	0.7	1	0.69	0.76	1.02	1.01	0.15	2.42	6.85	0.72	-1
C	1.33	1.23	1.22	1.31	1.27	0.79	0.46	0.46	0.76	0.62	1	0	0.95	0.83	1.22	0.24	2.47	10.29	0.46	1
D	1.05	1.26	1.26	1.05	1.15	0.91	0.76	0.76	0.91	0.84	1	0.92	0.76	1	1.01	0.11	2.07	6.46	0.83	0
E	1.64	1.44	1.45	1.67	1.55	0.24	0.24	0.13	0.13	0.19	1	0.96	0.76	0.24	1.24	0.38	2.97	13.83	1.17	1
F	1.69	1.31	1.78	1.15	1.48	0.72	0.1	0.03	0.05	0.22	1	0	0.66	0.71	0.45	0.41	2.88	11.62	0.57	1
G	1.35	1.32	1.24	1.41	1.33	0.1	0.68	0.68	0.1	0.39	1	1.12	1.17	0.89	1.23	0.31	3.21	13.31	1.08	1
H	1.55	1.55	1.57	1.57	1.56	0.24	0.24	0.14	0.14	0.19	1	0.97	0.97	0.24	0.13	0.38	3	12.79	5.33	1
I	1.63	1.63	1.63	1.63	1.63	0.3	0.3	0.3	0.3	0.3	1	0.29	0.29	1.61	1.61	0.46	2.1	7.66	0.18	1
J	1.83	0	1.37	1.14	1.08	0.07	47.82	0.56	0.08	12.13	1	0.6	0	0	1.14	0.42	2.38	9.13	0.53	1
K	1.7	1.45	1.46	1.74	1.59	0.44	0.34	0.08	0.09	0.24	1	0.51	0.84	0.51	0.08	0.39	2.85	11.22	2.27	1
L	0	1.75	1.13	1.67	1.14	44.47	0.05	0.04	0.6	11.29	1	0	0.53	0	1	0.43	2.06	9.07	0.53	1
M	1.65	1.64	1.66	1.66	1.65	0.16	0.16	0.12	0.13	0.14	1	0.52	1.15	0	1.2	0.4	4.05	17.99	1.39	1
N	1.6	1.59	1.58	1.59	1.59	0.23	0.26	0.19	0.28	0.24	1	0.99	0.97	0.47	0.37	0.39	3.41	13.5	2.35	1
O	1.04	1.04	1.04	1.04	1.04	0.94	0.94	0.94	0.94	0.94	1	0.96	0.94	1.02	1.01	0.03	2.03	5.79	0.94	0
P	1.19	1.19	1.58	1.05	1.25	0.95	0.72	0.51	0.51	0.67	1	1.09	0.72	0.95	0.51	0.26	2.47	7.54	1.24	0
Q	1.07	1.09	0.98	1.34	1.12	0.84	0.96	0.9	0.85	0.89	1	0.84	0.96	1.06	0.9	0.1	2.11	6.38	0.92	0
R	1.18	1.32	1.44	1.62	1.39	0.82	0.83	0.27	0.33	0.56	1	0.82	0.83	1.03	0.33	0.29	2.67	9.22	1.22	0
S	1.4	1.43	1.39	1.39	1.4	0.29	0.33	0.16	0.21	0.24	1	0.79	0.62	0.33	1.35	0.35	3.43	13	0.84	1
T	1.47	1.47	1.78	1.78	1.62	0.25	0.25	0.25	0.25	0.25	1	0.25	0.25	0.99	1.76	0.43	2.3	9.79	0.18	1
U	1.56	1.57	1.18	1.18	1.37	0.49	0.5	0.5	0.49	0.49	1	0.92	0.93	0	1.13	0.28	2.49	11.25	1.63	1
V	1.76	1.75	1.44	1.45	1.6	0.32	0.26	0.23	0.27	0.27	1	0.81	0.79	0	1.42	0.41	2.85	10.92	1.12	1
W	1.8	1.81	1.47	1.49	1.64	0.15	0.13	0.15	0.15	0.14	1	0.15	1.2	0.51	0	0.41	3.53	16.22	2.64	1
X	1.57	1.57	1.61	1.58	1.58	0.28	0.28	0.31	0.32	0.3	1	0.28	0.31	0.49	0.38	0.38	2.5	9.93	0.68	1
Y	1.65	1.66	1.66	1.66	1.66	0.08	0.05	0.47	0.02	0.15	1	0.56	0.57	0	1.64	0.44	2.45	9.4	0.69	1
Z	1.5	1.44	1.47	1.49	1.48	0.23	0.28	0.27	0.16	0.23	1	0.19	0.36	1.21	1.13	0.35	2.39	10.68	0.23	1
0	1.23	1.22	1.23	1.23	1.23	0.77	0.76	0.76	0.77	0.77	1	0.77	0.76	1.21	1.21	0.16	2.4	6.12	0.63	0
1	1.54	1.47	1.77	1.82	1.65	0.45	0.13	0.13	0.45	0.29	1	0.45	0.99	1.47	1.76	0.45	2.71	9.13	0.44	1
2	1.42	1.42	1.53	1.46	1.46	0.13	0.06	0.19	0.36	0.19	1	0.52	0.12	1.4	1.22	0.36	2.72	10.52	0.24	1
3	1.49	1.53	1.48	1.45	1.49	0.34	0.33	0.05	0.07	0.2	1	0.57	0.33	1.04	0.05	0.39	3	11.33	0.82	1
4	1.51	1.23	1.29	1.35	1.34	0.53	0.71	0.7	0.54	0.62	1	0.53	0.85	1.23	0.7	0.25	2.79	7.94	0.72	0
5	1.59	1.52	1.47	1.47	1.51	0.44	0.42	0.02	0.05	0.23	1	0.8	0.89	0.43	1.45	0.37	3.35	12.67	0.9	1
6	1.4	1.38	1.32	1.34	1.36	0.39	0.32	0.83	0.81	0.59	1	0.81	0.83	0.38	1.31	0.29	3.08	9.11	0.97	0
7	1.4	1.39	1.9	1.18	1.47	0.07	0.51	0.08	0.04	0.18	1	0.04	0	0.51	0.08	0.45	2.4	9.28	0.07	1
8	1.27	1.28	1.25	1.27	1.27	0.43	0.41	0.62	0.64	0.52	1	0.58	0.56	1.25	1.24	0.25	2.61	6.73	0.46	-1
9	1.32	1.34	1.41	1.37	1.36	0.83	0.8	0.38	0.32	0.58	1	0.83	0.8	1.32	0.38	0.29	3.09	9.11	0.97	0

REFERENCES

[1]. Sung-Bae cho "Neural-Network Classifiers for RecognizingTotallyUnconstrained Handwritten Numerals" IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 8, NO. 1, 2013

[2]. Swapnil Khedekar Vemulapati Ramanaprasad Srirangaraj Setlur, "Text - Image Separation in Devanagari Documents", Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR'03), 2003

[3]. Q. Yuan, C.L. Tan, "Text Extraction from Gray Scale Document Images Using Edge Information". Proceedings of the International Conference on Document Analysis and Recognition, ICDAR'01, Seattle, USA, pp. 302-306 September 10- 13, 2001

[4]. A. Antonacopoulos and R T Ritchings "Segmentation and Classification of Document Images", The Institution of Electrical Engineers, U.K. 1995.

[5]. G. Nagy, S. Seth, "Hierarchical Representation of Optically Scanned Documents". Proceedings of 7th Intl. Conf. on Pattern Recognition, Montreal, Canada, pp. 347-349, 1984.

[6]. Kyong-Ho Lee, Student Yoon-Chul Choy, and Sung-Bae Cho, "Geometric Structure Analysis of Document Images: A Knowledge-Based Approach", IEEE transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 11, November 2000.

[7]. Jean Duong, Hubert Emptoz, "Features for Printed Document Image Analysis", IEEE Transaction on Document Image Analysis Vol. 02, No. 17 November 2002.

[8]. Nikoluos G. Bourbakis, "A Methodology of Separating Images from Text Using an OCR Approach", Binghamton University, Center for Intelligent Systems, Binghamton, NY 13902.

[9]. Karl Tombre, Salvatore Tabbone, Loïc Pélissier, Bart Lamiroy, and Philippe, "Text/Graphics Separation Revisited" Proceedings of the 5th International Workshop on Document Analysis Systems, 2002.

[10]. V Chew Lim Tan, Zheng Zhang, "Text block segmentation using pyramid structure", Proceedings of SPIE, 4307, 297, 2000.