I.J. Image, Graphics and Signal Processing, **2015, 2, 8-14**
Published Online January 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2015.02.02

Density Based Script Identification of a Multilingual Document Image

Rumaan Bashir
Department of Computer Science, Islamic University of Science & Technology,
Awantipora, Pulwama, J&K, 192122, India
Email: rumaan.bashir@islamicuniversity.edu.in

S. M. K. Quadri
P.G. Department of Computer Science, University of Kashmir, Hazratbal, Srinagar, 190006, India
Email: quadrismk@hotmail.com

Abstract—Automatic Pattern Recognition field has witnessed enormous growth in the past few decades. Being an essential element of Pattern Recognition, Document Image Analysis is the procedure of analyzing a document image with the intention of working out the contents so that they can be manipulated as per the requirements at various levels. It involves various procedures like document classification, organizing, conversion, identification and many more. Since a document chiefly contains text, Script Identification has grown to be a very important area of this field. A Script comprises the text of a document or a manuscript. It is a scheme of written characters and symbols used to write a particular language. Languages are written using scripts, but script itself is made up of symbols. Every language has its own set of symbols used for writing it. Sometimes different languages are written using the same script, but with marginal modification. Script Identification has been performed for unilingual, bilingual and multilingual document images. But, negligible work has been reported for Kashmiri script. In this paper, we are analyzing and experimentally testing statistical approach for identification of Kashmiri script in a document image along with Roman, Devanagari & Urdu scripts. The identification is performed on offline machine-printed scripts and yields promising results.

Index Terms—Document Image Analysis, Multilingual Script Identification, Kashmiri, Roman, Devanagari, Urdu, Density, Statistical Approach.

I. Introduction

Computer Science has become inevitably a part of everything in the world that we encompass. It has limitless applications in various fields of life. All these fields of life have one common thing being used as a basic item of manipulation i.e. a document. Documents are used as storage, processing, detailing, writing & reading units. These documents were initially manipulated by hand leading to an unwieldy & cumbersome setup. As the role of computer science & computing grew, these documents were started to be manipulated by the use of electronic devices. Still the collection of hardcopy documents used to grow exponentially. These hardcopy documents again challenged the researchers to develop methods such that these documents and their contents could be processed electronically in an automated fashion. This led to the concept of Document Image Analysis. It is the procedure of analyzing a document image with the intention of extracting the contents so that they can be used according to the need of various stakeholders. A majority of these documents contain text in written form. This text needs to be analyzed and processed. This text is technically known as Script. A document may be written in one or more scripts. The script possesses a unique collection of symbols with which it is written. Languages, which are spoken, are written using these scripts. For example, the Roman (commonly known as English) language is written using the Roman script. Similarly, the Urdu & Kashmiri languages are written using modified Arabic script [1]. In addition to the above, identification refers to the action of identifying something or simply to establish the identity of something. To identify means to ascertain the origin, nature, or definitive characteristics. Sometimes it is even referred to associate or affiliate something closely with a class. Therefore, Script Identification refers to the identification of the script(s) a particular document has been written in. This process of script identification is needed to be performed electronically. Script Identification is an important area of research in the field of Pattern Analysis and Recognition (also known as Image Processing). The input to the procedure of script identification is a document image, which refers to the electronic (typically binary) image of a document which contains text written in one or more scripts.

In this paper, we are presenting the script identification of a multilingual-offline document which contains four scripts namely Roman (English), Devanagari, Urdu & more importantly, Kashmiri. Only machine-printed (typeset) scripts are being considered for identification. Extremely little work has been performed for the identification of Kashmiri script [2]. The motivation behind this piece of work is to *pioneer* the identification

of Kashmiri script and its most closely related scripts. The focus of this piece of work is the *optimization* of the document processing which takes place for Kashmiri & related scripts such that more documents could be processed in less amount of time. The technique proposed for identification is a simple statistical approach. The methodology presented here statistically analyzes the image of the document with a concept called as Density. The rest of the paper is organized into the following parts:

1. Literature review which discusses the recent advances in the area of script identification & provides a classification of various script identification techniques.
2. Methodology which provides a detailed structure of the method used to solve the identification problem along with the features used. It also lists the algorithm used for the purpose of calculation of density.
3. Findings/Discussion which presents the experimental results and illustration of the methodology used.
4. The advantages and the future scope are highlighted in conclusion section.

II. Literature Review

Since the last three decades tremendous work has been done in the area of script identification. Both foreign and domestic scripts have been identified using various techniques & strategies. In the foreign context research has been performed on English, Arabic [3], Chinese, Han & Roman [4], Japanese, Persian, Korean, Greek & Russian scripts. In the domestic context, techniques for identifying scripts like Malayalam, Kannada, Gurumukhi, Telegu, Devanagari/Hindi, Bengali, Urdu, Gujarati, Oriya, Tamil, etc have been proposed [5][6]. Various promising methods & techniques have been proposed. With regards to the Kashmiri script, a script of regional importance, extremely little work has been performed. Kashmiri language is one of the official languages in India [7]. It does not have its own script. It is written using other scripts with some modifications. Formally, the Perso-Arabic script is used to write Kashmiri language [1]. This paper is a significant step to initiate the identification of Kashmiri script. When written using Perso-Arabic, the Kashmiri script has certain specific features [2]. These features are listed below in Table 1.

Table 1. Features of Kashmiri script.

S. No	Feature	Particulars
1.	Writing/Reading order	Right to Left
2.	Diacritical marks	Heavy usage (pesh, zabbar, zaer, hamza, dots, tashdeed, etc.)
3.	Complexity	More Cursive than English & Devanagari
4.	Distinct Separation between lines, words & letters	Very Rare

Many techniques have been proposed to identify various scripts. In recent advances, the scripts are identified using morphological reconstruction & hole filling [8], water reservoir techniques [9], projection profiles [10], moments & run-length histogram [3], top & bottom profiles [11] [12], vertical & horizontal profiles [13], steerable pyramid features [14], texture features [15] & wavelet energy histogram. Entropy or energy evaluation has been done in a multilingual document image wherein column entropies are considered for the purpose of identification [16]. Structural features using 62 dimensional feature set have been used to perform script identification in printed Indian documents [17]. Local Binary Patterns (LBP) and a modified version of their orientation together with a least square support vector machine has been used for script identification purpose on newspapers and books [18]. PCA based character level script identification has been presented along with relative entropy and Euclidean distance parameters [19]. A 41 dimensional feature set has been prepared and classifiers namely Logistic Model Tree, Sequential Minimal optimization, Random Forest, RBF Network, Multi-layer Perceptron, LibLINEAR, Fuzzy unordered rule induction are applied on them for handwritten script identification [20]. The results of these classifiers are compared to discover the best classifier. Structural features have also been experimented for both handwritten and machine-printed scripts [21]. A fuzzy based approach for word-level script identification in low resolution images has been proposed using horizontal run statistics and wavelet features [22]. A connected component and support vector machine based procedure has been presented to distinguish between handwritten and machine-printed text in a document image for Bangla script [23]. Texture features have also been used to determine the script of document images based on the concept that text has a particular visual appearance and a distinct visual texture [24].

After the survey of the available literature [25], it was culminated that the procedure of script identification can have many forms & types.

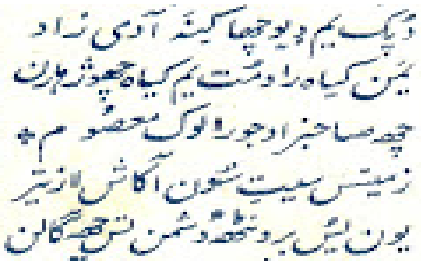


Fig 1: Example Kashmiri Script.

Based on this survey, a classification of script identification is given below:

A. *Method of Acquisition:*

i. Offline: This type of script identification is performed on a document which is already produced and not being produced.

ii. Online: This is the type of script identification wherein the script is being written & identified at the same instance of time. It is a real time system.

B. Method of Writing:

i. Handwritten: This type of script identification is applicable when the script has been written using hand and not any machine [26].
ii. Typeset: Here the script is printed using a machine e.g. a printer. Sometimes also called as machine-written.

C. Type of Script:

i. Foreign: This script identification involves the identification of scripts which do not belong to a specific country i.e. international scripts.
ii. Domestic: Here the scripts being identified are the scripts which belong to a specific country [27][9].

D. Features used for identification:

i. Local: When the feature used for identification fall within a small area of a document e.g. a word in a paragraph.
ii. Global: Here the features of the entire document as a single whole are taken into consideration for performing script identification.

E. Number of Scripts:

i. Unilingual: When a particular document under identification consists of only one script the identification is called as the unilingual script identification [28].
ii. Bilingual: When the document under identification consists of two scripts [2], the procedure is called as bilingual script identification.
iii. Multilingual: When a document under the procedure of identification consists of more than two scripts, the identification becomes complex, and is called multilingual script identification [27].

F. Technique of Identification:

i. Spatial Domain: When the pixels of an image are used as a base for identification the concept is called spatial domain. It has further two sub types:
   a. Statistical: Here the pixels are converted to statistical values are used for the identification purpose.
   b. Structural: Here the structure of the text & symbols delivered by the pixels is used to perform the identification [29].
ii. Frequency Domain: When an image is converted to frequencies and those frequencies are processed or manipulated to perform identification / classification it is called frequency-domain technique [30][31].

III. Methodology

An image is a two dimensional function f(x,y) where x and y are the spatial or Cartesian coordinates and f is the function representing the intensity values at location (x,y). Therefore, an image is essentially a collection of numerical values representing the intensity. These numerical values can be statistically manipulated to perform certain operations on the image. In order to perform the identification of Kashmiri script along with other related scripts viz. Roman, Urdu, & Devanagari, we propose a simple & statistical solution. Here statistical approach has been used to perform the identification of the four scripts using a concept called "Density". This Density is represented as Ø. Since the image is visually composed of small dots called pixels, the Density (Ø) here is the density of an image which is the count of the number of black pixels per unit of size used for calculation. It is a representative of how much dense the text part (written part) in a document is. It indicates the amount of fullness in a particular document and amount of emptiness on the contrary. This density is evaluated to check how much full the image is using the black part (or the written part) of it.

Initially, a document image is obtained from a valid source. This document image becomes the input to the procedure of script identification. This document image is then cropped to remove the unwanted portions (borders) in order to extract the text part only. This image is then divided into small pieces extracting lines from the text using a particular unit (in this case pixel). After dividing the image into constituent components, density of each part is calculated.

An image matrix which is the representative of the component of the image in the computer system becomes the base for the purpose of calculating the density. This image matrix is processed & manipulated. The illustration of this calculation is shown in the discussion section. The problem statement of this piece of work is given below:

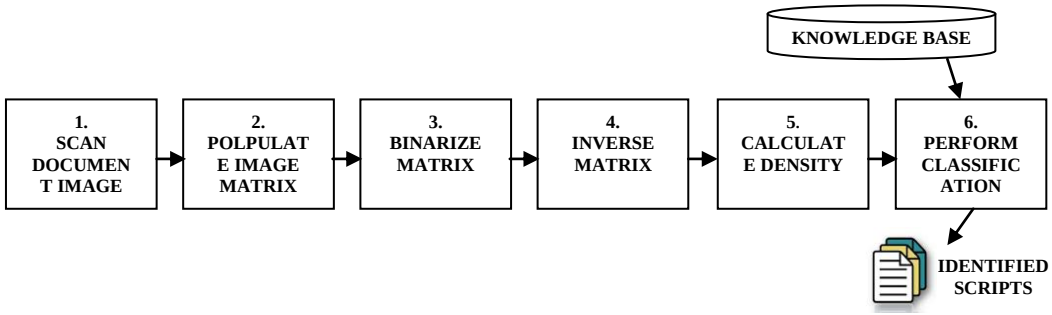


Fig 3. Script identification Methodology Used

Problem Statement:

To identify different scripts in an offline machine-written document image using statistical analysis.

Four scripts have been taken into consideration:

1. English / Roman, the globally important script
2. Hindi / Devanagari, the national script
3. Urdu, the formal script of Kashmir
4. Kashmiri, the local regional script

The algorithm to calculate the density of the image components is given below:

Algorithm ScriptIdnDensity():

Constraints: 1. Populated Knowledge Base for classification purpose. 2. Document image consists of texts only.

1. *Start*
2. *Scan the sample image*
3. *Populate image matrix in order to contain the intensity values*
4. *Binarize the image matrix*
5. *Invert the binarized matrix*
6. *Calculate the density of the matrix*
7. *Compare with the knowledge base*
8. *Perform classification*
9. *Stop*

Fig 2: Script Identification Algorithm Steps.

IV. Findings/Discussion

We have applied the proposed technique on a variety of printed document images. The script identification has been performed in two stages. The first stage is a training stage where in a set of 400 image components of document images were used for training the algorithm in order to create the knowledge base. After the creation of the knowledge base, the actual phase of script identification was commenced in which 800 image components of document images were used to test the efficacy of the algorithm.

All of these images were scanned from printed documents containing text in normalized form. These document image components considered for the experiment were of size 610 X 25 pixels representing a single line of unique script.

The document is scanned using a simple multipurpose scanner with a 300 dpi criterion. The document image looks like the image given in Fig. 4. Next, this document image is segmented into components of size 610 X 25 pixels using a segmentation algorithm in order to yield lines of unique scripts. This is shown in Fig. 5.

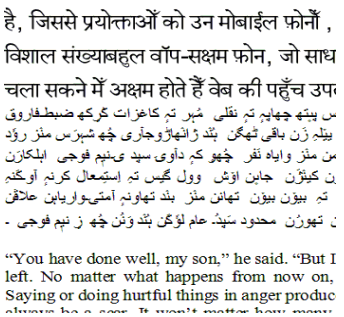
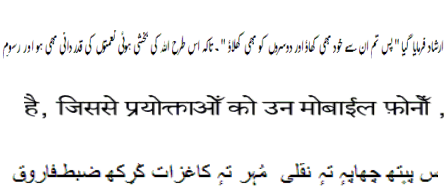

है, जिससे प्रयोक्ताओं को उन मोबाईल फ़ोनों ,
विशाल संख्याबहुल वॉप-सक्षम फ़ोन, जो साध
चला सकने में अक्षम होते हैं वेब की पहुँच उप
"You have done well, my son," he said. "But I
left. No matter what happens from now on,
Saying or doing hurtful things in anger produc

Fig 4: A scanned document image.


"You have done well, my son," he said. "But I

ارشاد فرمایا گیا "پس تم ان سے خود بھی کھاؤ اور دوسروں کو بھی کھلاؤ"۔ تاکہ اس طرح اللہ کی بخشی ہوئی نعمتوں کی قدردانی بھی ہو اور رسوم

है, जिससे प्रयोक्ताओं को उन मोबाईल फ़ोनों ,

س پیتھ چھاپِ تمِ نقلی مُہر تمِ کاغزات گرکھ ضبطِفاروق

Fig 5: Image components (script lines) with a unique script viz. Roman, Urdu, Devanagari & Kashmiri (in top to bottom order).

These individual components are then represented as image matrices as shown in the Fig. 6. This image matrix is actually the representation of the image in the computer. The image intensities are normally stored in a 2 dimensional data structure in rows and columns. Here, the image matrix is populated with the intensity values of which the image is comprised.

0	1	0	10	0	255	254	233
233	1	0	1	10	9	255	250
2	255	250	247	1	255	0	0
255	1	1	0	0	0	244	244
0	0	255	252	250	0	1	0
0	0	0	0	255	0	0	1

Fig 6: Image Matrix.

This matrix contains values ranging from 0 to 255. These values are there because the document image components are in gray scale. Since, the document is usually black & white therefore the values are near to 0 (for black) or near to 255 (for white). This image matrix is then converted to a binary matrix by approximating the 0's & adjacent values to 0 and approximating the 255 & adjacent values to 1. This step is called as binarization and helps to achieve a numerical value which can be used for further calculation. This binary matrix is shown in Fig. 7.

0	0	0	0	0	1	1	1
1	1	0	0	0	0	1	1
0	1	1	1	0	1	0	0
1	0	0	0	0	0	1	1
0	0	1	1	1	0	0	0
0	0	0	0	1	0	0	1

Fig 7: Binarized Matrix.

This binarized matrix is then inverted to convert 0 to 1 and 1 to 0 so as to indicate black as 1 (presence of content as true) and white as 0 (absence of content as false). Simply said, a complement of the binarized matrix is calculated. This complemented matrix is shown in Fig. 8 and is called the inverted matrix.

1	1	1	1	1	0	0	0
0	0	1	1	1	1	0	0
1	0	0	0	1	0	1	1
0	1	1	1	1	1	0	0
1	1	0	0	0	1	1	1
1	1	1	1	0	1	1	0

Fig 8: Inverted Matrix.

After the creation of inverted matrix, the Density is calculated as per equation (1).

$$\text{Density } (Ø) = \sum_{i=0}^{n} x_i \qquad (1)$$

where n is the number of cells in the matrix and x_i = {0,1}.

In the Fig. 8, the Density (Ø) is calculated to 29 and the total cells are 48. These values are converted to the total density peak percentage of (29/48) * 100 = 60.41 %. After evaluating the density, the script is to be classified as per the density factor. In the training stage, the outcome of which is revealed in Table 2, a set of images was used to train the algorithm. In the training phase, the knowledge base was created and the classification/identification index was made. This is shown below:

If Peak percentage for (Ø) == 4.288 % ±0.3% then
Script = Hindi
Else if Peak percentage for (Ø) == 4.911 % ±0.3% then
Script = Kashmiri
Else if Peak percentage for (Ø) == 2.862 % ±0.3% then
Script = English
Else if Peak percentage for (Ø) == 3.521 % ±0.3% then
Script = Urdu
Else
Script = Others

Table 2: Results of the Training Phase.

S. No	Script	Size of Image (pixels)	Density (Ø) Range	Peak percentage for (Ø)
1	Devanagari	610 X 25	613 to 695	4.288 %
2	Kashmiri	610 X 25	676 to 822	4.911 %
3	Roman	610 X 25	400 to 473	2.862 %
4	Urdu	610 X 25	494 to 580	3.521 %

In the testing stage 200 image components were tested for each script type and the results are listed in Table 3. The results of the proposed technique are highly significant and yield 97% overall accuracy.

Table 3: Final Results.

S. No	Type of Script	Total No. of Images Tested	Correctly Identified	Accuracy Rate
1	Roman	200	194	97%
2	Devanagari	200	196	98%
3	Kashmiri	200	194	97%
4	Urdu	200	192	96%
	Overall Accuracy Rate	800	776	97%

V. Conclusion

Document image analysis is a highly important area of research with direct application and implementation in the practical world with Script identification being an important procedure in it. Script identification has been performed in detail for many scripts. This work is a pioneering step for the identification of Kashmiri script together with its closely related scripts. Till date extremely little work has been reported for the identification of Kashmiri script.

The technique of density calculation has been proposed and experimentally tested for identifying Kashmiri and its related scripts. Density is used as a statistical feature for identification which indicates the fullness (written part) of a document image or the amount of content present. The technique proposed in this paper provides promising results. It is very simple to implement and gives concrete output. Further, it does not require any specialized equipment for identification purposes.

The work presented in this paper can be improved and extended in many ways some of which have been listed as under:

1. Improve the technique / method to increase the rate of recognition / identification. (fast processing & less memory i.e. optimize)
2. Extend the work presented to include more scripts.
3. Extend the work on handwritten Kashmiri script.
4. The technique could be applied to work out numerals of various scripts.
5. Typically, identify & experiment other features to distinguish between Urdu & Kashmiri as these are very similar scripts and very common scripts in Kashmir.

6. The concept of density could be used to solve or improve other issues of document image analysis.

REFERENCES

[1] S. S. Toshkhani, "Kashmiri Language: Roots, Evolution and Affinity". Kashmiri Overseas Association, Inc. (2008)

[2] Rumaan Bashir and Smk Quadri, "Identifiction of Kashmiri Script in a Bilingual Document Image", In Proc. 2nd IEEE ICIIP, JUIT Shimla, India, (2013).

[3] Ahmed M Elgammal & Mohamed A. Ismail, "Techniques for Language Identification for Hybrid Arabic-English Document Images", IEEE, (2001).

[4] Sukalpa Chanda, Umapada Pal, Katrin Franke & Fumitaka Kimura, "Script Identification- A Han and Roman Script Perspective", IEEE, Intl. Conf. on Pattern Recognition, (2010).

[5] Shamita Ghosh and Bidyut B. Chaudhuri, "Composite Script Identification and Orientation Detection for Indian Text Images", IEEE Intl. Conf. on Document Analysis and Recognition, (2011).

[6] Sukalpa Chanda, Umapada Pal & Katrin Franke, "Font Identification – In Context of Indic Script", 21st Intl. Conf. on Pattern Recognition (ICPR-2012), Japan.

[7] U. Pal, S. Sinha and B. B. Chaudhari, "Multi-Script identification from Indian documents", IEEE, Proc. 7th Intl. Conf. on Document Analysis and Recognition, (ICDAR 2003) vol 2. pp 880-884.

[8] B. V. Dhandra, P. Nagabushan, Mallikarjun Hangarge, Ravindra Hegadi, V.S. Malemath, "Script Identification Based on Morphological Reconstruction in Document images", IEEE, 18th Intl. Conf. on Pattern Recognition, (2006).

[9] U. Pal & B. B. Chaudhari, "Automatic Identification of English, Chinese, Arabic, Devanagari and Bangla Script Line", IEEE, (2001).

[10] M. C. Padma, P. A. Vijaya, P. Nagabushan, "Language Identification from an Indian Multilingual Document Using profile features", IEEE, Intl. Conf on Computer and Automation Engineering, (2009).

[11] M Swamy Das, D. Sandhya Rani, C. R. K. Reddy, " Heuristic based Script Identification from Multilingual Text Documents", IEEE, 1st Intl. Conf on Recent Advances in Information Technoogy (RAIT), (2012).

[12] P. A. Vijaya & M. C. Padma, "Text Line Identification from a Multilingual Document", IEEE, Intl. Conf. on Digital Image Processing, (2009).

[13] Prakash K. Aithal, Rajesh G., Dinesh U. Acharya, Krisnamoorthi M. Subbareddy N.V, "Text Line Script Identification for a Trilingual Document", IEEE, 2nd Intl. Conf. on Computing, Communication and Networking Technologies, (2010).

[14] Mohamed Benjelil, Remy Mullot, Adel M. Alimi, "Language and Script identification based on Steerable Pyramid Features", IEEE, Intl. Conf. on Frontiers in Handwriting Recognition, (2012).

[15] Hiremath P. S., Shivashankar S., Jagdeesh D Pujari & V. Mouneswara, "Script Identification in a handwritten document image using texture features", IEEE, Proc. 2nd Intl. Advance Computing Conf. (2010).

[16] Bashir, R. and Quadri, S.M.K., "Entropy based Script Identification of a multilingual Document Image", IEEE Intl. Conf. Computing for Sustainable Global Development (INDIACom), 2014 Page(s): 19 – 23.

[17] Obaidullah, S.M.; Mondal, A.; Roy, K. "Structural feature based approach for script identification from printed Indian document", IEEE Intl. Conf. on Signal Processing and Integrated Networks (SPIN), 2014 Page(s): 120 - 124

[18] Ferrer, M.A.; Morales, A.; Pal, U., "LBP Based Line-Wise Script Identification", 12th IEEE Intl. Conf. on Document Analysis and Recognition (ICDAR), 2013 Page(s): 369 – 373.

[19] Mingji Piao; Rongyi Cui, "An Approach to Script Identification in Multi-language Text Image", IEEE 6th Intl. Conf. on Intelligent Networks and Intelligent Systems (ICINIS), 2013 Page(s): 248 – 251.

[20] Obaidullah, S.M.; Roy, K.; Das, N., "Comparison of different classifiers for script identificationfrom handwritten document", IEEE Intl. Conf. on Signal Processing, Computing and Control (ISPCC), 2013 Page(s): 1 – 6.

[21] Saidani, A.; Echi, A.K.; Belaid, A. , "Identification of Machine-Printed and Handwritten Words in Arabic and Latin Scripts", IEEE 12th Intl. Conf. on Document Analysis and Recognition (ICDAR), 2013 Page(s): 798 – 802.

[22] Angadi, S.A.; Kodabagi, M.M., "A fuzzy approach for word level script identification of text in low resolution display board images using wavelet features", Intl. Conf. on Advances in Computing, Communications and Informatics (ICACCI), 2013 Page(s): 1804 – 1811.

[23] Banerjee, P.; Chaudhuri, B.B., "A System for Handwritten and Machine-Printed Text Separation in Bangla Document Images", Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR), 2012 Page(s): 758 – 762.

[24] Busch, Andrew; Boles, W.W.; Sridharan, S., "Texture for script identification", Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume: 27, Issue: 11 (2005), Page(s): 1720 – 1732.

[25] Debashis Ghosh, Tulika Dube, & Adamane P. Shivprasad, "Script Recognition – A Review", IEEE, Trans. On PAMI Vol. 32 No. 12 pp 2142-2161 (2010).

[26] K. Roy, S. Kundu Das, Sk Md Obaidullah, "Script Identification from Handwritten Documents", IEEE Proc. 3rd Intl. Conf on Computer Vision, Pattern Recognition, Image Processing and Graphics, (2011).

[27] U. Pal & B. B. Chaudhari, "Script Line Separation From Indian Multilingual Script Documents", Proc 5th Intl. Conf, on Document Analysis ad Recognition, IEEE Comp. Society Press pp 406-409 (1999).

[28] Faoud Slimane, Slim Kanoun, Jean Hennebert, Adel M. Alimi, Rolf Ingold, "A study on font family and font size recognition applied to Arabic word images at ultra-low resolution", Elsevier, Pattern Recognition Letters 34 (2013), 209-218.

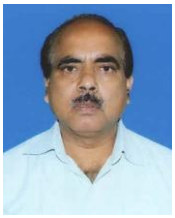
[29] Rajesh Gopakumar, N. V. Subareddy, Krishnamoorthi Makkithaya & U. Dinesh Acharya, "Zone-based Structural feature extraction for Script Identification from Indian Documents", IEEE, 5th Intl. Conf. on Industrial & Information Systems, (2010).

[30] Zhou, L.; Ping, X.J.; Zheng, E.G.; Guo, L., "Script identification based on wavelet energy histogram moment features", 10th IEEE Intl. Conf. on Signal Processing (ICSP), 2010 Page(s): 980 – 983.

[31] Jianjia Pan; Yuanyan Tang, "A rotation-robust script identification based on BEMD and LBP", IEEE Intl. Conf. on Wavelet Analysis and Pattern Recognition (ICWAPR), 2011 Page(s): 165 – 170.

Author's Profiles

Ms. Rumaan Bashir is working as Assistant Professor in the Department of Computer Science, Islamic University of Science and Technology, Awantipora, J&K, India. Her area of research is Pattern Recognition/Document Image Analysis and is presently working on Script Identification. The author received her MCA and PGDBA from the University of Kashmir, J&K, India. She has done M.Phil in Computer Science and is currently pursuing her Ph. D from University of Kashmir, J&K, India.



Prof. (Dr.) S.M.K. Quadri is working as Head, Department of Computer Science, University of Kashmir, Srinagar, J&K, India and Director, Directorate of Information Technology & Support System, University of Kashmir, Srinagar, J&K, India. His research areas include Software Reliability, Programming Language, Simulation and Modeling, Software Engineering & Pattern Recognition. He received his M.Sc in Statistics and PG Diploma in Computer Programming, both from Aligarh Muslim University, India. He also received his M.Tech from Indian School of Mines, Dhanbad and his Ph. D (Computer Science) from University of Kashmir, J&K India.

How to cite this paper: Rumaan Bashir, S. M. K. Quadri,"Density Based Script Identification of a Multilingual Document Image", IJIGSP, vol.7, no.2, pp.8-14, 2015.DOI: 10.5815/ijigsp.2015.02.02

