International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2
***Article Received:** 05 January 2022 **Revised:** 28 January 2022 **Accepted:** 20 February 2022*

Identification of Kashmiri Script in a Document Image Using Dwt and Entropy

Rumaan Bashir[1], Kaiser J. Giri[2], Javaid Iqbal Bhat[3]*, Zahid Hussain Wani[4]
[1]Department of Computer Science, Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-rumaan.bashir@islamicuniversity.edu.in
[2]Department of Computer Science Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-kaiser.giri@islamicuniversity.edu.in
[3*]Department of Computer Science Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-javaidonnet@gmail.com
[4]Department of Computer Science Islamic University of Science & Technology, Awantipora, J & K (India)
Email:-zahid.uok@gmail.com
***Corresponding Author:-** Javaid Iqbal Bhat
*Email:-javaidonnet@gmail.com

Abstract:
Over the past decades, the growth in the development of automated systems and their related technologies for document processing has been exponential to achieve more and more effective & efficient solutions. Since the majority of documents contain text so text processing is an important aspect of the document processing. However for text processing it is important to know in which script a particular document is written. These scripts are used to write languages. The identification of script has been an important area of research and accordingly a substantial amount of research work has been done using different schemes. However, there is very less work reported with respect to the identification of Kashmiri language. The main impetus in this work is to perform the identification of Kashmiri Script alongside Urdu, Devanagari and English. The scheme uses discrete wavelet transformation and entropy evaluation. The results achieved are promising. This work will also pave a way for designing more optimal and better solutions with respect to overall automated solutions to the Kashmiri Language.

Keywords: Automatic Document Processing, Script Identification, Discrete Wavelet Transformation, Entropy.

1 Introduction

Describing & depicting the inner self to the world has always required an arduous effort from man since the olden times. However, humans have discovered numerous ways for this depiction and are even said to have accomplished the ability. Expression of thoughts has been one of the important achievements over the past centuries in the history of human race. In this, spoken expression has taken the first place but written expression is not in any way behind. Writing has become the key and noteworthy manner of the depiction [1]. Writing is a pictorial and perceptible practice of representing language.

Human communication that represents language & thoughts through the inscription or recording of signs & symbols is called 'Writing'. Writing schemes use sets of symbols to depict the sounds of speech. It may also have symbols for punctuation & numerals [2]. As human civilizations progressed, the development of writing was driven by practical necessities such as sharing information, recording keeping, marketing activities, etc. Around the 4th millennium BCE [3], the complexity of government &

trade in Mesopotamia outgrew human recall, and writing became a more reliable system of recording & presenting transactions in a perpetual form. In both Mesoamerica & Ancient Egypt, writing systems may have evolved through calendric events and a political requirement for recording important history.

With the innovation of computer based systems and more recently communication systems, a duplication of the writing systems has transpired [4]. Computer systems are used to write on & read from as we are used to writing on and reading from paper. The electronic documents are the electronic form of items

which replace the paper. Electronic documents which are comprised of these writings are created, stored & processed in these computers in diverse languages [5]. Today, hundreds of different languages are used by the computer systems. Scripts are used to write languages and there and many scripts used in the world. The same scripts are used to write in the computers. Script is the assortment of characters & alphabets used for writing [6] [7]. This script has now become a crucial & essential aspect of electronic documents. Before any text processing could be applied one needs to identify the script in which the electronic document has been written. Due to this very fact, past decades have seen a great deal of research and development in Automatic Document Processing [8].

The earliest practical OCR, to which identification of script is antecedent, appeared in the 1950's in the United States. This was the same period wherein the UNIVAC (preliminary commercial computer) [9] was made. There have been significant developments in the OCR technology since then.

The automation process & huge document processing has increased the total amount of hardcopy and softcopy documents. Fields of Cloud, Grid, & Big Data Computing have also augmented the quantity of information on the internet. Therefore storage, retrieval, classification and manipulation of such documents has become more interesting day by day. This has led to the development of the area of Document Image Analysis wherein in the documents are scanned as images and then processed for data extraction [6] [7] [10].

As the information revolution has taken over the whole world, there are some realms remain unexplored like the identification of Kashmiri script in a document image, the local native regional script, in the Kashmir Valley located in the Jammu & Kashmir, India. Koshur as it is called locally is a language belongs to the Indo-Aryan language's Dardic sub-group [10]. For the development of the Kashmiri language, since November 2008 it has been made a compulsory subject in schools in the Kashmir Valley. Kashmiri is the native script of Jammu & Kashmir and is closely written alongside Roman (English) – the international script, Devanagari (Hindi) – the national script & Urdu – the local script. Kashmiri language is written using perso-arabic script and is a highly phonetic language. In the context of a multilingual system, a document may contain a combination of any of these four scripts. Apropos the above, script identification needs to be performed for Kashmiri script and the other related scripts present in a document image.

In the current scenario, many script identification schemes have been designed for different scripts all over the entire orb, however the attempt to identify Kashmiri script has started a few years back [6][7][10]. The work reported performs spatial-domain script identification. However, frequency domain is known to provide wide and better solutions with respect to image processing. Therefore, the prime motivation behind this research work is an endeavor to contribute for the development & establishment of automatic, computer-based solutions for Kashmiri script –"the local native regional script" and thus the language in terms of Frequency Domain

2 Literature Review

The Automatic Document Processing has been one of the primary field of research over the past few decades and processing of documents in image form has also been one of the most active areas of research [8]. This Document Image Analysis involves lot of aspects like graphical/textual, colour/grayscale, 2D/3D etc. In textual document image analysis, the identification of scripts is mandatory for automatic document image analysis and numerous systems have been designed & many scripts, international & national, stand identified. Out of these, Roman (English) script has been at the centre stage. Script identification focuses on identifying the script of the language in which a document image is written. The identification is easy if the document image is written in one script i.e. unilingual. The processing & identification of scripts becomes complex when the document image contains more than one script owing to the fact that each script has distinctive features that helps distinguish them e.g. English & Hindi are different since in Hindi all text lines have a top bar called as sirorekha [11] [12] which is absent in latter. Moreover, some scripts share features e.g. Urdu and Arabic use nearly same symbols. Some other factors such as the directionality of writing also help in identification. Some scripts like English are written left to right, while Urdu is written right to left. Others are written horizontally or vertically like Chinese & Japanese scripts.

2.1 Script identification using frequency domain:

The document image in the frequency domain, is transformed to a set of frequencies. These set of frequencies are operated in order to perform the

identification of scripts. A survey conducted for automatic document processing wherein the systems for identifying the scripts have been presented [13]. A comprehensive analysis of script identification, its approaches highlighting various Indian scripts has been presented [14]. A thorough survey on identification of Indic Scripts [15] describing types, features, classifiers has been performed. Rotation-invariant texture features based identification of scripts is [16] given for Greek, Chinese, English, Malayalam, Persian & Russian scripts. Biologically inspired texture-based technique for English, Japanese, Hebrew, Indonesian, Russian, Persian, Korean, Oriya, Malayalam & Kannada scripts for text analysis has been presented using local energy [17]. Texture theory based script identification [18] [19] for different scripts using diverse texture features has been performed. For Chinese, Arabic, Korean and Hindi scripts, a multi-classifier system [20] consisting of GMM, Nearest Neighbour, SVM, and Euclidean Distance is implemented. Steerable Gabor Filters have been used on Chinese, Korean, Japanese and English [21]. A BEMD and LBP based rotation robust script identfiication has been given [22]. A word-level multi-script identification different font styles and sizes has also been performed [23]. The scripts tested using this include Roman, Kannada, Devanagari, Bengali, Gujarati, Malayalam, Oriya, Tamil, Telegu, Urdu & Punjabi. Here, the efficacy of Discrete Cosine Transforms and Gabor features has been independently assessed using SVM and NN. Texture classification has been given [24] using wavelet based co-occurrence histogram features for English, Hindi, Kannada, Bengali, Telegu, Tamil, Urdu & Malayalam. The texture features are taken out using the correlation between subbands at the same resolution. A handwritten script identification has also been presented [25]. Roman, Devanagari & Bangla scripts have been subjected to a two-stage word wise script identification [26]. Indian scripts including Bangla, Gurumukhi, Malayalam Oriya, Devanagari, Telegu & Roman for multi-script document images has been performed word-level identification of scripts using texture based features like Moment invariants and Histograms of Oriented Gradients (HOG) [27]. Here, MLP found to be a better classifier. Block level handwritten script identification for English, Malayalam, Kannada, Hindi, Tamil, Telegu & using Discrete Curvelet Transforms are presented [28]. Here, DCvT mines directional selective features efficiently in comparison to Discrete Wavelet Transform since directional discriminating properties such as curves, lines & edges it are not highlighted. Handwritten Indic script identification on the basis of Convolution based technique [10] has been given for Bangla, Roman, Urdu, and Devanagari. Gabor filter & Morphological reconstruction which are convolution based methods are combined to construct a feature vector of 20 dimensions. MLP is used for classification. Arabic & Latin script identification on document images, both handwritten & machine written are presented using steerable pyramid features [29]. Here, the input image is initially fragmented into two sub-bands; a high-pass and low-pass using filters. Computed mean, energy, homogeneity, standard deviation, kurtosis, etc. features are selected. The overall results and outcomes attained through these methods are significant.

2.2 Research Gap:

The already available works in this area have concentrated mainly on the scripts used in writing languages like Chinese, Roman, Arabic, Korean, Russian, Cyrillic, Han, Hebrew, Sinhala, Persian, Thai in the international context and Devanagari, Kannada, Tamil, Malayalam, Gurumukhi, Oriya, Bangla, Gujarati, etc. in the domestic context. The work related to identification of Kashmiri script in document images is reported in [6] [7] [10], though through spatial domain. The frequency domain for identification of Kashmiri script has not been explored in detail. This work shows the script identification of Kashmiri Language in document images in frequency domain. Therefore, the research gaps can be enlisted as under:

1. Most of the work related to this field focuses on the identification of other official scripts in India.
2. The work reported for script identification of Kashmiri in a document image is experimented & implemented in spatial domain mostly.

3 Research Work

3.1 Theory:

This study revolves around two main concepts. The Discrete Wavelet Transformation & Image Entropy. The Discrete Wavelet Transform (DWT) is one among the frequently applied frequency-domain transforms [30]. The other common frequency domain transforms include the Discrete Cosine Transform (DCT) and Discrete Fourier Transform (DFT). DWT has been found effective across a range of application in image processing. Here, an image is decomposed into various

constituent frequencies. This is following by the selection of the most suitable portion for further processing. DWT has been used in digital image processing since it possesses superb spatial localization & multi-resolution features and because of this it imitates the theoretical models of human visual system [31]. Discrete Wavelet Transformation has proven to be effective with regard processing in the frequency domain [32] [33] especially because of hierarchical image decomposition features [34]. Wavelets comprise of small waves of multi-resolution analysis & restricted duration are non-stationary in nature and are better processed using the application of DWT [35]. In addition to the retention of temporal information, wavelet transformation provides spatial as well as frequency representation of an image [36] [37] [38] [39]. This feature is not available in Fourier Transformation (FT). Single-level 2-D wavelet decomposition is performed with the help of DWT. It is performed with respect to either a certain wavelet or certain wavelet filters. First DWT is applied to segregate a signal into its high frequency & low frequency components. Commonly, these hold edge information (high frequency) & smooth variations (low frequency). Discrete Wavelet Transformation up to Single-level with HAAR filter was applied on all images. Here, the approximation coefficients matrix & detailed coefficients matrices of each input image are computed. The selection of the frequency of interest is performed which provides a compact representation of the original image. This is done to improvise and optimize the following steps.

Image Entropy, on the other hand, is the degree of variation in an image. This is an indication of the amount of energy in energy present in the image. It is an index of the activity of a portion of an image in comparison to other portions [6]. The evaluation of entropy in mathematical terms is performed by the calulation of the number of transitions. Image Binarization is performed to arrive at the matrix which contains only two values so that entropy calculation can be performed. For images which are represented in binary form, the transition from a 0 to 1 and 1 to 0 are the indicators of entropy. These transitions are also recorded as positive if the binary value changes from 0 to 1 and negative if a binary value changes from 1 to 0 in a particular data set [40]. As shown in the table below, entropy values can be categorised on different parameters.

Table 1: Type of Entropies

S. No	Type	Particulars
1	Positive Entropy	Transition from 0 to 1
2	Negative Entropy	Transition from 1 to 0
3	Local Entropy	Transitions in Portion of Image
4	Global Entropy	Transitions in Whole Image
5	Column Entropy	Transitions in column of an Image Matrix
6	Row Entropy	Transitions in row of an Image Matrix

Different images possess different entropy values. Therefore, this becomes a highly possible factor for their differentiation. The entropy values obtained for a training set are used later to classify the test data set successfully. Here, the column entropies both negative & positive are used for calculation of the final entropy.

3.2 Classification:

The classification of script identification through frequency domain is proposed as given in the figure below. The classification is performed on various parameters like the acquisition method, writing method, features used, colour information and number of scripts. The identification of script in frequency domain can be performed in an image which has been produced as a document already and needs to be scanned for the identification purpose. An image which is being identified for script alongside when the script is being written is called as online script identification. In frequency domain, the features used for script identification as to whether local (part of image) or global (whole image) determines the type of script identification. The frequency domain identification of scripts can be further of two types depending upon whether the image under consideration is a black & white or colour image. The number of scripts present also determines the type of identification.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2
Article Received: *05 January 2022* ***Revised:*** *28 January 2022* ***Accepted:*** *20 February 2022*

3.3 Dataset Used:

Since sufficient dataset of Kashmiri script is not available, we have been motivated to prepare a fresh data set of document images for script document written in
Kashmiri, Devanagari, English & Urdu. Therefore, a dataset using available machinery was prepared. For the experiment of our proposed system, we considered good quality document images of different machine written scripts in A4 format. A total of 400 images were prepared. These document images contained English, Devanagari, Kashmiri and Urdu scripts. These images normally use a 8-bit bitmap file format (.bmp). Black and white text documents were considered with text only wherein the text was written in single column format. The identification algorithms use an entire script line with running text for identification purposes. The images of the document are scanned at a minimum 300 dpi in grey-scale. Each script line image has a size of 35 X 640 pixels (common A4 text).

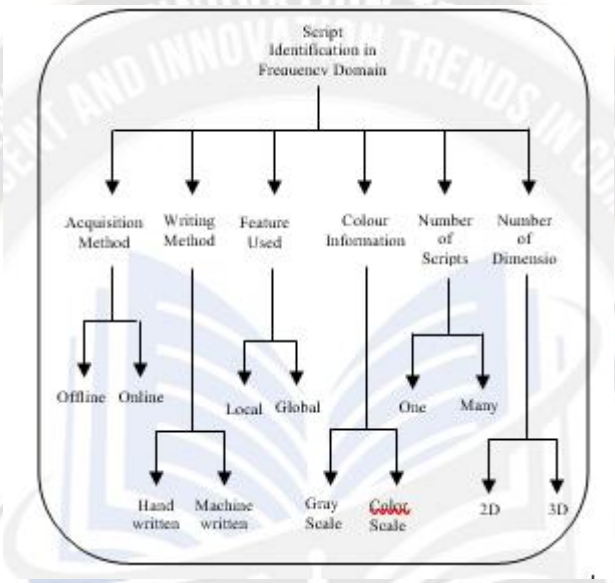

Figure 1: Classification for Script Identification in Frequency Domain

3.4 Methodology:

In The empirical & experimental research methodology has been applied in order extract results from real world implementations. There are two phases of implementation.

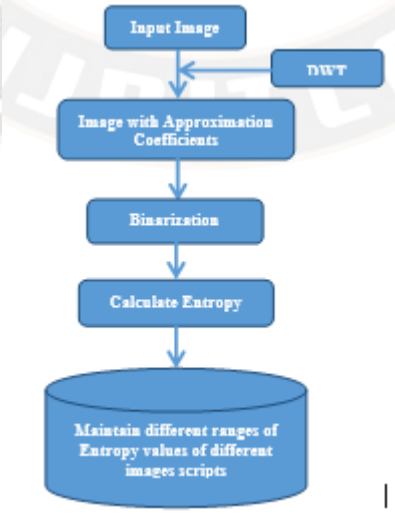

Figure 2: Training phase used for Script Identification.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2
Article Received: *05 January 2022* ***Revised:*** *28 January 2022* ***Accepted:*** *20 February 2022*

The training phase which generated the necessary database for entropy values and the testing phase which actually checks the image for the type of script.

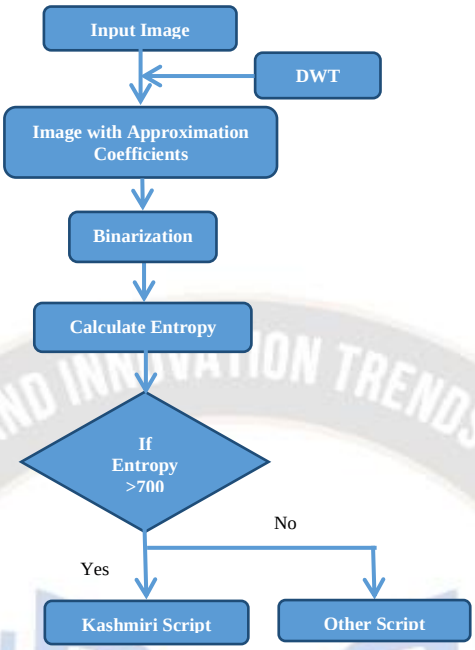

Figure 3: Testing phase used for Script Identification.

The general method involves a six-step procedure which is depicted in the figure below.

Fig 4: General Model used for Script Identification.

These steps are explained as under:

Step 1:

Acquisition of image is done using a general purpose flat-bed scanner. The black and white image is scanned. The image is having a white background on which the text is written using one script in black color.

Step 2:

The preprocessing of the image is done for noise removal, skew correction, and line segmentation. This is done by using already available algorithms. The outcome of this step is a collection of images in which a single line of script is available.

Step 3:

As discussed above these images one by one are subjected to the DWT. The application of DWT is followed by the selection of most suitable frequency component. The HAAR filter is used to perform the DWT. In our proposed model, out of all the matrices generated by DWT, we selected the approximation coefficients matrix for further processing. Each input image yields a respective coefficients matrix. It is important to mention that the size of this representation of the image is far less than the original size of the image.

Step 4:

When the image matrix of the suitable to the requirement is identified using DWT as given above, these matrices are subjected to binarization. Binarization is performed for converting image matrix into a binary matrix which contains only 0 and 1 values. So, all the images of all scripts from the newly generated database containing approximation coefficient matrices are binarized.

Step 5:

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2
***Article Received:** 05 January 2022 **Revised:** 28 January 2022 **Accepted:** 20 February 2022*

After this binarization, the entropy is calculated in the image matrix. For the purpose of entropy calculation only column entropy both positive and negative is calculated.

Step 6:

The entropy calculated in the above step is matched with the entropy values generated earlier in the training phase. This matching step allows classifying the image containing the Kashmiri script from other images containing other three scripts.

3.5 Experimental Work Done

In line with the proposed work plan and objectives outlined, some of the existing models and state of art to work with the identification of any particular script among various document images of different scripts were researched. Basic models and their mathematical theory was explored. Besides understanding these mathematical models, existing practical implementation was understood and verified. Experiments can test the accuracy of theories. It is important to restate that all the experiments and results should be reproducible. A lot of experiments were carried out to arrive at the final outcome. The dataset was prepared, algorithm was designed and analyzed. The experiment has been conducted and the obtained results were tabulated followed by their analysis.

3.6 Experimental Analysis

The proposed Kashmiri script identification scheme was carried by performing two independent tests during the experimental phase. A sample set of four (04) different image document scripts namely English, Kashmiri, Urdu and Devanagari were selected for experimental purpose. The initial experiments, direct entropy calculation was performed after acquisition, preprocessing and binarization of all the images of all scripts. This was done to have an insight about the entropy calculations ane their respective scales. The following experiments were conducted as per the model and method given above. It is pertinent to mention that in the DWT/Entropy method, the number of matrices of different scripts generated remains the same as that of the original database, however, the dimension of each image in newly generated database reduces to one fourth i.e.320 x 18 pixels as compared to original image dimension of 640 x 35 pixels thus improvising & optimising the direct entropy calculations. This was followed by the Binarization of all the image matrices followed by the calculation of entropy in each of the binarized matrices. It is again mentioned that the entropy was calculated in column fashion however both negative as well as positive. The results obtained using this experimentation easily classify the images as into Kashmiri and Non-Kashmri scripts based on the calculated entropy values.

Table 2 below shows the experimental analysis results for above mentioned tests. For the description purpose, results of only three images from each script are presented here. From the obtained results it is evident that the column entropy values of none of the document image scripts after applying DWT transformation goes beyond 590 except Kashmiri scripts which is above 700 for all of its images. So using this noticeable entropy difference between Kashmiri script images and other remaining document image scripts, a well result oriented classification has been carried out.

Scripts	**Images**		**Physical Images**	**Column Entropy**	**Execution Time**
English Script	Image 1	Original	your co-authors must be members of this website and while submitting the paper you	2518	0.000790
		DWT Transformed		434	0.000389
	Image 2	Original	your co-authors must be members of this website and while submitting the paper you	2452	0.000485
		DWT Transformed		514	0.000251
	Image 3	Original	paper - You can now submit your paper using the Submit Paper link. Please note that all	2446	0.000486
		DWT Transformed		522	0.000671

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2
***Article Received:** 05 January 2022 **Revised:** 28 January 2022 **Accepted:** 20 February 2022*

Kashmiri Script	Image 1	Original	تہٕ وومیزٕ ؤسی ؤسی پوان چھےٚ پمن پتیٚمہٕ ؤری یہٕ نوٚ سرٕ کٔن لوٚگمٕت اوس ۔ عوٚآمی خلقن منٛز پمن لتھن	2410	0.000800
		DWT Transformed	[illegible]	**724**	0.000398
	Image 2	Original	تہٕ دون مُلکن درمیان تازٕ بیان بآزی سام ببوان چھوٚننہٕ پوان ز دوشوے مُلک ہِندُستان تہٕ پآکِستان چھہٕ آکھ	2394	0.000542
		DWT Transformed	[illegible]	**728**	0.000201
	Image 3	Original	نوٚو روخ پانناوٕتھ بیان بآزی گران تہٕ کآشرِین سِتی ہمدردی وٚرتاوٕتھ سٕے پزون ماحول ووتلاوان ییٚمہٕ کِنی	2478	0.000495
		DWT Transformed	[illegible]	**704**	0.000183
Urdu Script	Image 1	Original	حج کی دنیوی برکات ومنارشاد فرمایا گیا کہ تاکہ یہ سب حاضر ہوں اپنے طرح طرح کے فائدوں کیلئے جو کہ دنیوی بھی ہیں اور اخروی بھی ۔ دنیوی	2334	0.000766
		DWT Transformed	[illegible]	492	0.000401
	Image 2	Original	یہ کہ اس میں طرح طرح کی تجارت کانفع کاذکر : سوحج کی دنیاوی برکات ومنافع کے ذکر کے طور پر بڑے بڑے مواقع ہوتے ہیں جو کہ بالا	2270	0.000518
		DWT Transformed	[illegible]	486	0.000197
	Image 3	Original	جماع جائز ہے ۔ جبکہ تجارت کو مقصود اصلی نہ بنایا جائے ۔ بلکہ یہ اس لحاظ سے مستحسن بھی ہے کہ اس سے یہ ظاہر ہوتا ہے کہ اسلام میں	2102	0.000492
		DWT Transformed	[illegible]	572	0.000253
Hindi Script	Image 1	Original	पर्यावरण से वास्तव में एक कागज रहित इलेक्ट्रॉनिक पर्यावरण के लिए एक गंभीर संक्रमण की	2566	0.000739
		DWT Transformed	[illegible]	482	0.000404
	Image 2	Original	गई है कि देखा है. यह विकास और प्रौद्योगिकियों कंप्यूटिंग और संचार का व्यापक उपयोग की	2640	0.000484
		DWT Transformed	[illegible]	488	0.000210
	Image 3	Original	भारी मात्रा के कारण हुई है इस के अलावा स्वत दस्तावेज़ छवि विश्लेषण दरियादिली से इस	2600	0.000483
		DWT Transformed	[illegible]	446	0.000194

Table 2: Sample scripts, their DWT transformation & respecive entropy values

3.7 Illustration

The illustration of above experiment is given below and is depicted in figure 5. The scanning of the document is using a common scanner with a 300 dpi yielding a document image as shown in figure 5. After basic pre-processing, the segmentation of the document image is performed. The image is segmented into smaller components with size of 640 X 35 pixels. This generated single lines of text with unique scripts. These individual components of scripts are next converted to their equivalent approximation coefficient image matrices after applying DWT transformation using HAAR filter which are represented as image matrices and contain values from 0 to 255 owning to the grayscale nature of the document image components. The document being usually black & white which makes the values are near to 0 for black or near to 255 for white. This image matrix obtained here is then transformed to a binary matrix. This is done by approximating the 0's & values near to 0 to 0 and approximating the 255 & values near to 255 to 1. Subsequently calculation of entropy is done. Here, Column entropy was used which includes both positive as well as negative transitions across each column for calculating the entropy, the script is to be classified as per the entropy value. The training phase includes the creation of the knowledge base and the classification/identification index. After conducting the training phase, a set of images were tested for efficacy

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2
***Article Received:** 05 January 2022 **Revised:** 28 January 2022 **Accepted:** 20 February 2022*

and the result of the proposed technique is listed in the table 3.

In the testing stage, 100 line script images of each script were tested and the results are presented in table 3. The results are promising with an accuracy of 98.25%.

Table 3: Experimental Result.

S. No.	Type of Script	No of Script Lines Tested	No of Script line correctly identified
1.	Kashmiri	100	99
2.	English	100	97
3.	Urdu	100	98
4.	Devanagari	100	99
	Total	400	393

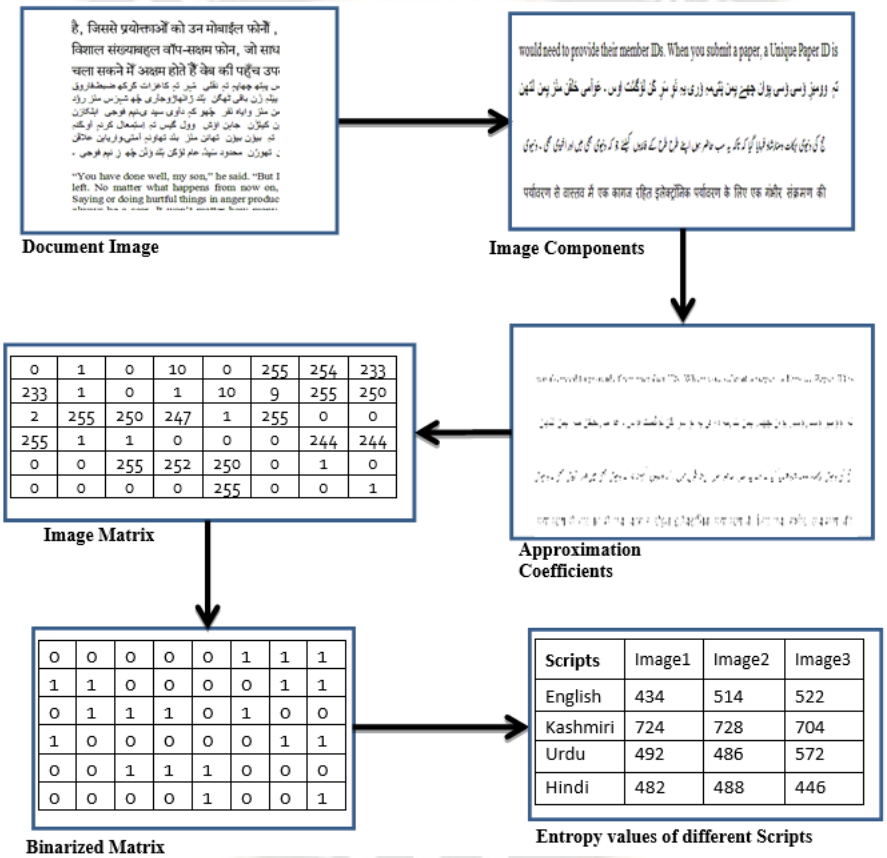

Figure 5: Illustration of the experiment.

Conclusion and Future Scope

The work presented in this paper is associated with the identification of scripts in automatic electronic document processing. Four major scripts have been taken into consideration which includes Kashmiri, English, Devanagari and Urdu. This piece of work focuses on Kashmiri script being the native language of Kashmir. The method used is based on Discrete Wavelet Transformation followed by calculation of Image Entropies. The method involves a six step procedure which includes image acquisition, preprocessing, application of DWT, binarization, entropy calculation and recognition. The experimental work shows promising results. The experiment successfully identifies / differentiates the four scripts. The DWT reduces the processing to a smaller matrix achieving improvisation and optimization. The method proposed in this paper can be used to:

- Recognize scripts which are written online dynamically.

- Recognize 3D fashion of scripts.
- Recognize coloured documents.

References

[1] . (1922). A Short History Of The World.
[2] Coulmas, Florian. 2003. Writing systems. An introduction. Cambridge University Press.
[3] Clinton Robinson, October 2003, UNESCO
[4] Rumaan Bashir, S. M. K. Quadri and Kaiser Javeed, "Script identification: a Review", 2018, vol-10, Page-1-15 International Journal of Information Technology/Springer Singapore.
[5] Coulmas, Florian (1996). The Blackwell Encyclopedia of Writing Systems. Oxford: Blackwell Publishers Ltd. ISBN 0-631-21481-X.
[6] Bashir, R., & Quadri, S. M. K. (2014, March). Entropy based script identification of a multilingual document image. In 2014 International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 19-23). IEEE.
[7] Bashir, R., & Quadri, S. M. K. (2015). Density based script identification of a multilingual document image. Int J Image Graph Signal Process, 2, 8-14.
[8] Rumaan Bashir, Kaiser J. Giri., "A Study of Script Identification Techniques", (2016), Volume (21), Pages(173-186) Research Cell: International Journal of Engineering Sciences.
[9] Hiromichi Fujisawa, "Forty years of research in character and document recognition- an industrial perspective", Elsevier Pattern Recognition 41 (2008) 2435-2446.
[10] Giri, K. J., & Bashir, R. (2013, December). Character recognition based on structural analysis using code & decision matrix. In 2013 International Conference on Machine Intelligence and Research Advancement (pp. 450-453). IEEE.
[11] Sk Md Obaidullah, Nibaran Das and Kaushik Roy, "Convolution Based Technique for Indic Script Identification from Handwritten Document Images", I. J. Image, Graphics and Signal Processing, 2015, 5, 49-57.
[12] B. V. Dhandra, H. Mallikarjun, Ravindra Hegadi and V. S. Malemath, "Word-wise Script Indentification from Bilingual Document Based on Morphological Reconstruction", IEEE 2006.
[13] Yuan Y. Tang, Seong-Whan Lee and Ching Y.Suen, "Automatic Document Processing: A Survey", Elsevier, Pattern Recognition, Vol. 20, No. 12, pp. 1931-1952 (1996).
[14] Debashis Ghosh, Tulika Dube, & Adamane P. Shivprasad, "Script Recognition – A Review", IEEE, Trans. On PAMI Vol. 32 No. 12 pp 2142-2161 (2010).
[15] Pawan Kumar Singh, Ram Sarkar and Mita Nasipuri, "Offline Script Identification from multilingual Indic-script documents: A state-of-the-art", Elsevier Computer Science Review 15-16 (2015)1-28.
[16] T. N. Tan, "Rotation Invariant Texture Featues and Their Use in Automatic Script Indentification", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 20, No. 20, July 1998.
[17] Woei Chan, George Coghill, "Text analysis using local energy", Elsevier Pattern Recognition, 34, (2001) 2523-2532.
[18] 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.
[19] A. Busch, "Multi-Font Script Identififcation Using Texture-Based Features", In Proc. Intl. Conf. Image Analysis and Recognition, pp 844-852, Sept 2006.
[20] S. Jaeger, H. Ma and D. Doermann, "Identifying Script on Word-Level with Informational Confidence", In Proc. Intl. Conf. Document Analysis and Recognition, Vol. 1, pp. 416-420, Aug. 2005.
[21] W. M. Pan, C. Y. Suen and T. D. Bui, "Script Identification Using Steerable Gabor Filters", In Proc. 8th Intl. Conf. on Document Analysis and Recognition , 2005
[22] 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.
[23] Peeta Basa Pati and A. G. Ramakrishnan, "Word Level multi-script identification", Elsevier Pattern Recognition Letters 29 (2008) 1218-1229.
[24] 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).

[25] P. S. Hiremath and S. Shivashankar, "Wavelet based co-occurrence histogram features for texture classification with an application to script identification in a document image", Elsevier Pattern Recognition Letters 29 (2008) 1182-1189.

[26] Sukalpa Chanda, Srikanta Pal, Katrin Franke and Umapada Pal, "Two-stage Approach for Word-wise Script Identification", In Proc. IEEE 10th Intl. Conf. on Document Analsysis and Recognition, 2009.

[27] Pawan Kumar Singh, Ram Sarkar and Mita Nasipuri, "Word-level Script Identification Using Texture Based Features", International Journal of System Dynamics Applications 4(2) 74-94 Apr-Jun 2015.

[28] B. V. Dhandra, Vijaylaxmi M. B. And Mallikarjun Hangarge, "Script Identification using Discrete Curvelet Transforms", International Journal of Computer Recent Advances in Information Technology, 2014.

[29] 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).

[30] K. J. Giri and R. Bashir, "Digital Watermarking: A Potential Solution for Multimedia Authentication," Studies in Computational Intelligence, pp. 93–112, Oct. 2016.

[31] K. J. Giri, R. Bashir, and J. I. Bhat, "A Discrete Wavelet Based Watermarking Scheme for Authentication of Medical Images," International Journal of E-Health and Medical Communications, vol. 10, no. 4, pp. 30–38, Oct. 2019.

[32] Kumar A, "A Review on Implementation of Digital Image Watermarking Techniques Using LSB and DWT", In: Information and Communication Technology for Sustainable Development. Springer, Singapore, pp 595–602 (2020)

[33] Kaiser J. Giri, S. M. K. Quadri, Rumaan Bashir & Javaid Iqbal Bhat "DWT based color image watermarking: a review", Multimedia Tools and Applications An International Journal ISSN 1380-7501 Volume 79 Combined 43-44 (2020)

[34] Liu K-C, "Human visual system based watermarking for color images", Fifth Int Confer Information Assurance Secur 2:623–626(2009)

[35] Qiang S, Hongbin Z, "Color image self-embedding and watermarking based on DWT", Int Conf Meas Technol Mechatron Autom 1:796–799(2010).

[36] Giri, K. J., Quadri, S. M. K., Bashir, R., & Bhat, J. I. (2020). DWT based color image watermarking:a review. Multimedia Tools and Applications, 79(43), 32881-32895.

[37] Giri, K. J., Bashir, R., & Bhat, J. I. (2019). A discrete wavelet based watermarking scheme for authentication of medical images. International Journal of E-Health and Medical Communications (IJEHMC), 10(4), 30-38.

[38] Giri, K. J., & Bashir, R. (2018). A block based watermarking approach for color images using discrete wavelet transformation. International Journal of Information Technology, 10(2), 139-146.

[39] Giri, K. J., & Bashir, R. (2017). Digital watermarking: a potential solution for multimedia authentication. In Intelligent techniques in signal processing for multimedia security (pp. 93-112). Springer, Cham.

[40] Gowda, S.D. & Nagabhushan, P., "Entropy Quantifiers Useful for Establishing Equivalence between Text Document Images", Conference on Computational Intelligence and Multimedia Applications, Intl. Conf.on Volume:3,(2007) , pp(s): 420 - 425