Texture Features from Handwritten Images for Writer Identification

Rashmi Mundas
Mtech In Information Technology, ISE Department
SDM College Of Engg & Tech., Dharwad
KARNATAKA, INDIA
rashmism232@gmail.com

Deepa Bendegeri
Assistant Professor, ISE Deparment
SDM College Of Engg & Tech., Dharwad
KARNATAKA, INDIA
deepabendigeri@yahoo.co.in

Dr. Jagadeesh Pujari
HOD, ISE Deparment
SDM College Of Engg & Tech., Dharwad
KARNATAKA, INDIA
jaggudp@yahoo.com

***Abstract*—** Identification of the writer is having wide scope in emerging technology due to its usage in various types of applications, especially in forensic science and biometric science. Our aim in this project is to identify author or writer from script which is handwritten and obtained as scanned images. Features of textures will be elicitated from wavelet decomposed images based on co-occurrence histograms. These will get (capture) the information about the relations among sub-bands of less frequency and that in sub-bands of higher frequency at the particular level of the transformed image. If the co-relation between the sub-bands has resolution of same then that indicates a stronger relation. Then relationship strength will indicate as information was essential considered to differentiating the textures. The proposed methodology will be executed with English handwritten images by considering 5, 10 penmanship or writers. Ability of features from texture in identifying writers is indicated though the outcome achieved in experimentation.

Keywords-*Wavelet; Texture features;Writer identification; document image; scanned images; co occurrence histograms*

*****

I. INTRODUCTION

Writer identification is a framework designed for recognizing the writer of a manually written record. An arrangement of reports from known authors must be known ahead of time to select another archive to one of this writer. To begin with, components (features) are registered on the writer of a reference report and after that these feature parts will be appeared differently in relation to the ones which exist in the database set. An author having most noteworthy closeness to the existing is related to archive. Writer identification distinguishes the handwriting methodology by taking into account of unknown handwriting image which continues by coordinating obscure handwritten image against a database of composed examples with known writer. In this manner, writer identification is imperative with numerous applications, for example, report examination, security, and monetary action, measurable and utilized as access control. The difficulties for identification of the writer and retrieving the author incorporates from various pen features, that differs in the writer handwriting style of their composition, if certainty which the author (writer) has composed content in the rush or not, furthermore that single word is uncommon composed the very same way twice.

Writer identification technique falls in classifications: Text dependent and text independent. Text-dependent approaches require handwritten tests taking into account on a particular content, or expect handwriting recognizer accessible for checking realness of writer. Writer identification utilizing signature is most prominent occurrence of these sorts of methodologies. Text-dependent approaches have advantage that they utilize the information of the substance of the information to isolated style from substance. This will build the precision of text-dependent frameworks. The significant issue of text-dependent frameworks is that they are non-relevant to situations when the content is not accessible, for example, in criminal equity frameworks includes examination between content archives with various substance. Also, message subordinate frameworks are much inclined to falsification when same information is displayed during testing. These sorts of frameworks can be actualized in the agreeable environment, where significant concern is exactness and writer might be asked to author particular substance to demonstrate their personality. These strategies are fundamentally the same as signature check procedures which includes the correlation between individual characters or words which have known semantic substance. Subsequently these techniques require earlier restriction and division of the right data, which is generally performed by human client collaboration.

The text-independent writer identification framework displays the style data, free of the content, which is utilized to recognize the writer in view of any given text content. This requires the insights of features which are computed from large extensive quantity of data to avoid anomalies due to specific content. Writer identification and verification technique utilizes measurable elements which are extracted from entire image containing a text segment (block). Base measure of manual written (e.g, few content lines in section) is extremely important to determine stable features components which will be hard to the text content of the examples. Consequently our technique will fall in this later class.

Writer identification includes input of two types: on-line and off-line. Online technique includes catching of pen development of author, where the style is compelled. Online writer identification framework uses temporal succession (sequence) code, which tracks pressure and velocity(speed) varies in handwriting, and pattern(shape codes) that relay on direction of trajectory in writing was developed for Chinese and English language [Bangy Li and Tieniu Tan 2009]. It works better for little number of characters. Online text-independent writer recognition framework [PitakThunswarin and TakenobuMatssura] for the language Thai depends on speed of pen pointer utilizes Fourier change technique.

In offline writer identification framework, scanned image of the author composing is utilized which delineates his behaviour. Offline text independent writer identification using Hidden Markov Model [Andreas Schlapbach and Horst Bunke] works on the basis of computing the score unknown author and comparing it scores of every individual author. The score of every individual author is computed by recognizer in view of hand writing. The recognizer with the most astounding score is assigned as unknown author. In offline writer identification framework, the hand written text of the writer is filtered and utilized for feature extraction. In that capacity offline writer identifications postures more difficulties contrasted with on-line method on account of the absence of extra features, which are accessible to online frameworks, is absent for offline system. Statistical based Writer identification method for non consistently skewed handwriting images has been discussed in [H S Said, K Baker and also T N Tan]. Different strategies for writer identification depends on Contour based features, Hierarchical Shape Primitive elements, has been talked about in [Mohamed NidhalAbdi and Maher Khemakhem]. Offline text-independent writer identification is very imperative for measurable examination, archives approval, and calligraphic relics ID.

II. FEATURE EXTRACTION

A. Discrete wavelet transform

The continuous 1-D transform wavelet of (1D) signal F(m) described as

$(WaF)(b) = \int F(x)\psi^*_{a,b}(x)dx$

ψ represents wavelet calculated according to Base ψ wavelet interpretation, expansion

$\psi_{x,y}(M) = 1/\sqrt{|a|}\ \psi\ (x-a/b)$

Due to certain conditions, base wavelet ψ fulfills limitation which includes zero mean. It is differentiated through control of x,y to distinct cross section ($x=2^y$, $b \in l$). Normally, forced it changes to non excess, integrated and that it should include multi-resolution primitive signal view. Expansion of Two-D is generally operates utilizing the result of One-D wavelet channels. The Haar wavelet is defined as

$\Psi(t) = \{1\ \ -1\ \ 0 \qquad 0<=t<=1/2,\ ½<=t<1,\ \text{otherwise}$

Above equation is much easier, where it checks $\{\psi_{n,a} = (2^{-n/2}\ \psi\ (2^{-n/2}\ s - a))\}_{n,a\in Z}$ is orthogonal and unit vectors premise for $L^2(R)$. In this, discretization $x=2^a$ and $y=s2^a$ is utilized, which will be p sought after all through this area. This wavelet is, verifiably, primarily identified wavelet.

The simplest way to compute a 2D discrete wavelet transform (DWT) of an image is to apply one-dimensional transform over image rows and columns separately and then to carry out down sampling. This transform decomposes an image with the overall scale factor of four, providing at each level one low resolution subimage and three wavelet coefficient subimages

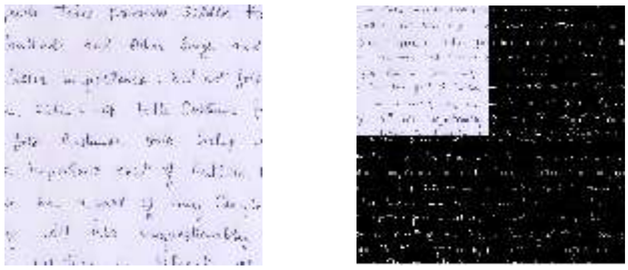
$$A = |S_x * | S_y * I |_{2,1}|_{2,1}$$

$$H = |T_x * | S_y * I |_{2,1}|_{2,1}$$

$$V = |S_x * | T_y * I |_{2,1}|_{2,1}$$

$$D = |G_x * | G_y * I |_{2,1}|_{2,1}$$

Here I is the input image. Sx, Sy and Tx, Ty represent low and high pass filters respectively, * denotes the convolution operator and 2 denotes downsampling operation. The subbands labeled H, V, D correspond to the detail images, while A corresponds to the approximation image as shown in the Fig. 1.



(a) (b)

Fig 1 (a) Original image
(b) 1-level wavelet transformed image (Haar)

B. Computation scheme

Script or style types usually vary from each other by the way they are assembled or grouped into words, and also the state of individual attribute, and so forth. This gives diverse scripts particularly distinctive visual appearance. Texture can be characterized in simple definitive form as "same pattern occurring repetitively" or something comprising of commonly related elements. This identification of script or writer style from the handwritten images consist features which are based on texture, extracted from handwritten images provided by writer in English Language The feature extraction method is described below.

The computation scheme is extraction of features influenced by perception of individuals is talented for recognition among new writings simply in view of simple visualization analysis. Classification of texture is processed by considering the Identification of script. Hence, this is complicated visible texture made out derived by sub-pattern. Despite of fact that, the sub-patterns can have scarcity of a better mathematical standard, it is well entrenched that a texture is considered as completely only if all the sub-patterns are correctly defined. We utilize a multi-resolution method as elicitation of texture features according to DWT and by using minimum distance classifier, the classification of the textures is obtained. This extraction of features is illustrated as:

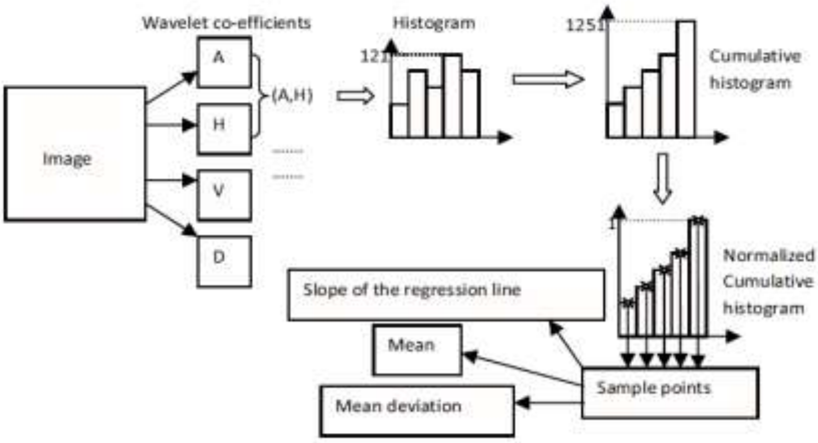


Fig. 2. Schematic diagram of the feature extraction algorithm

Handwritten image is considered as input image X and using Haar wavelet we apply 2D discrete wavelet transform (DWT), which provides us approximation sub-band image (I), and also detail sub-band images (H,V,D) (Fig 3.3). We take set of two images like (I, V), calculate the co occurrence histograms H1, H2 in provided direction. For every histogram, normalized cumulationis composed for all co-occurrence histogram and then enumerate the texture features, especially, mean, regression-line slope and mean-deviation described as above. The overall procedure will be looped in eight directions which yields two-histograms * three-features * eight-directions = 48 components features from each set (I, V).

Correspondingly, for all sets such as, (I,H), (I,D), (I, abs(V-H-D)) by elicitation of features, we obtain 192 components to given handwritten image (I).

The detailed process will be repeated for the image which is complement image of I, represented as $[\bar{I}] = [255 -$ where i is gray-value pixel for image I. From I and features extracted for combining and obtaining a feature space containing dimension of 384. These are used for training of features and then classification. Fig 3.3 illustrates the feature extraction schema in detail. The schematic diagram of the feature extraction method is shown in the Fig. 2.

III. TEXTURE TRAINING AND CLASSIFICATION

A. Training

Training step includes extraction of features from different handwritten image samples, which are chosen randomly that belongs to every script using above feature extraction methodology. These extracted features will be saved in the feature library. Then these will be utilized for writer identification.

B. Classification

Classification step involves comparison of features values with that saved in feature library. The extraction of features for image I is done using feature extraction method, as explained above. Then using distance vector code these will be compared with subsequent component values which will be saved in library,

D (N) = $\sqrt{\sum_{l=1}^{M}[f_l}$

Herein, N represents overall components in component space f, f_l (Y) expresses jth component of texture of given example X, and f_l (N) reproduces lth texture value for Mth class in library. After this process, the written record is determined by minimum distance classifier.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Writer identification have effective approach in image document analysis, this observation is considered for grouping of composition. This approach is discussed in our project; efficiency of feature elicitation method for texture is articulated. We already discussed about different writers, who have different sense of writing in different state of mind, which also includes different styles. Hence we consider, a text block as distinct pattern. This study helps us to motivate the use of texture classification for identification of writer. Our approach will not involve connected component method. Since, this is a global approach used for texture classification. We perform experiments by allowing the authors to write different scripts in English language. These scripts are digitized each to 150 dpi. Execution of our method is done on 10 different writers. From each writer we have taken 10 images, from which, we considered 5 images for training, other 5 images for testing.

Fig 3 : Sample images of English Script written by different writers

	Case 1	Case 2	Average
256*256	95.0	97.5	96.25
512*512	97.5	100	98.75
1024*1024	92.5	95.0	93.75

Table 1: Average classification accuracies(%) for different size.

V. CONCLUSION

Here, we implemented a technique for extraction of texture features from handwritten images for identification of writer. Texture features will be extracted from wavelet decomposed images by the use of the relationship between the sub-bands, on the basis of co occurrence histograms. This methodology results as a powerful differentiator among the authors or writers comparing various manually written scanned images. The executed results demonstrate the capacity from proposed technique furthermore capability of the wide worldwide methodology to distinguish proof of author in examination of record/report investigation which includes wide significance from the scientific as well as biometric discipline.

REFERENCES

[1] P.S. Hiremath, Shivashankar S, Jagdeesh D Pujari, Mouneswara., "Script Identification in a handwritten document image using texture features", In Proc. IEEE 2nd International Advance Computing Conference, Patiala, pp. 110-114, 2010.
[2] Said, H.E.S, T N Tan and K D Baker., "Personal identification based on handwriting", Pattern Recognition, 2000.
[3] Tan T., "Rotation Invariant Texture Features and their Use in Automatic Script identification", IEEE Trans. Pattern Analysis and Machine Intelligence, pp. 751-756.
[4] Zhu, Y., Tan, Wang Y., "Font Recognition Based on Global Texture Analysis", IEEE Trans, Pattern Analysis and Machine Intelligence, vol 23, no 10,pp.1192-1200.
[5] Zois, E N and V Anastassopoulos., "Morphological waveform coding for writer identification", Pattern Recognition 385-398, 2000.
[6] Hertel and Marti., "A set of novel features for writer identification in Audio and Video Based Biometric Person Authentication", ed. J Kittler and M S Nixon, pp. 679-687, 2003.
[7] Schlapbach A., Bunke., "A Writer identification and verification system using HMM based recognizers", Pattern Analysis Application (Springer).