International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 90-94
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

OFFLINE TEXT-INDEPENDENT WRITER IDENTIFICATION

Anupama Inamdar
PG Student,
Department of Computer Science and Engg
B.L.D.E.A CET, Vijayapur,
Karnataka, India

Abstract—**Handwriting identification technique is the basic need for criminological examination, record approval, calligraphic relics identification. It works in online and offline categories. The proposed automatic offline text-independent writer identification method works in three stages such as training, enrollment, and identifications. In all the stages firstly, the handwriting image is segmented into word regions (WRs) using an isotropic Laplacian of Gaussian (LoG) filter. Then the features extracted from the each word regions by using important feature extraction techniques such as Scale Invariant Feature Transform (SIFT) and Gray Level Co-occurrence Matrix (GLCM). Comparing to GLCM, the SIFT method is more suitable and efficient to identify the writer because it extract the structural features of handwritten text. The proposed method is tested on real time data collected of 42 samples (6 writers, 7 samples from each writer). By the experimental result, the proposed method strongly suggest that SIFT features are efficient in handwritten text identification.**

Keywords— **Text-based image searching, Reranking Image Retrieving, Hypergraph.**

I. INTRODUCTION

In general, writing styles, shapes, and characters are different from one person to another. Finding out the authorship of a questioned document could provide valuable information for an investigation this process is called writer identification. This identification is works on offline text and this has several applications they are criminological examination, records approval and calligraphic relic's identification etc. [1].

Handwriting identification technique is divided into online and offline categories. The offline handwriting identification includes text dependent technique and text independent technique. Therefore, writer identification recently has been studied and it has a wide variety of applications, such as security, financial activity, forensic and used as access control. Handwriting is a personal biometric that find unique persons [2].

As a rule, on-line frameworks accomplish better execution since they can utilize spatial and transient data about the composition. Logged off frameworks, then again, are hard to outline subsequent to the dynamic data about the composition is lost amid the report obtaining. They likewise can be classified into content ward and content free. As the name recommends, in the content ward strategies, the composition tests contain a predefined content [3].

Broad researches have been led in this field. Furthermore, progressions of universal author recognizable proof challenges [4, 5] have been effectively sorted out. There will be two types of texts; they are text-dependent which means the writing samples are predefined one and another one is text-independent which means any text to establish the writer identity.

The samples of authors or writers depending upon the situation they are writing and the mood of the writers. Plamondon et al. [6] exhibited a thorough overview of early research writings as for programmed essayist distinguishing proof. In early days the identification or investigation is done manually which takes more time and sometimes the given result is not accurate. But today all work is done automatically using new technologies and advanced methodologies. For identifying hand writing images of writer or a person there will be mainly two approaches texture-based approaches and Structured-based approaches.

Texture means repeated pattern of pixels and structure means something that is strong enough for construction. Writer identification works on both text dependent and text independent. Textures based methodologies take hand writings as an uncommon composition picture and concentrate the textural highlights for identification.

Hanusiak et al. [7], H Said et al. [8] and Zhu et al. [9] utilized a method of "grey level co-occurrence matrix (GLCM)" to getting the intensity values from the hand writing pictures. To remove the unwanted noise and clarity purpose "Hidden Markov Tree (HMT)" model is used for hand writing identification.

Helli et al. [10] and He et al. [11] use new methods for getting better results. They use a "wavelet-based" texture based approach. Du et al [12] use a "Gabor and XGabor" method to identifying handwriting images collected from different authors. Along with this method "feature relation graph (FRG)" method is used for automatic hand writing identification.

Bertolini et al. [13] proposes new methods for identification and verification of handwriting images. "Local binary pattern

(LBP)" and "local phase quantization (LPQ)" are the two texture-methods for identification and verification.

To improve the offline text-independent writer identification the structure based approaches are more intuitionistic, prominent and stable. In early time structure based approaches are more popular.

Bulacu et al. [14] mainly work on the edges of handwriting contents using "edge-based directional" method and identifying the unique person. Here the features are extracted based on the edges.

The rest of the paper is organized as follows. Literature survey in section II. Proposed algorithm in section III. Experimental results are presented in section IV. Concluding remarks are given in section V.

II. LITERATURE SURVEY

Schomaker et al. [15] employed the coordinate of points of handwriting image and these points collected to form feature vector table. Firstly they normalize and fragment connected component and these components are more in values. Means these values create confusion for the system, for this purpose the codebook is generated for storing calculated values. Codebook generation is good habit for identifying images. In processing stage large set of writers are used for handwriting identification.

Bulacu et al. [14] proposed to characterize the individuality of writer's using "edge-based directional feature" extraction method. Edge based means it consist of edge direction distribution and edge-hinge distribution. To find the writers individuality they proposed effective and new techniques for writer identification. Here they use texture based approaches for defining property of methods of the handwritten samples.

Helli et al. [10] proposed "Gabor and X Gabor filter" and "feature relation graph (FRG)" methods. The first method extracts features from handwriting content of images and the second method is used represent the extracted features. Using Gabor and X Gabor filter method the writer identification is done. Here the filters are used to improve the images of handwriting. Feature relation graph is constructed for every handwriting image for identification.

Bertolini et al. [13] proposed a method for identification and verification of handwriting images. Here they use both "local binary patterns (LBP)" and "local phase quantization (LPQ)" method. The "LBP and LPQ" apply on handwriting images to extract textural features of given images. The given image is converted into binary images and calculates local key points. The quantization is applied to given images.

Jain et al. [2] proposed a method of" K-adjacent segment means (KAS)" for writer identification. Here the standard databases are used to test the proposed method and results are analyzed. In this method codebook is generated for mean values and the identification is done comparing with trained images.

Djeddi et al. [12] proposed a new method of identifying writers, which is concentrated on connected components. Here the image segmented then codebook is generated to identify the writers. With fixed window size the image is segmented and features are extracted from segmented images. Codebook features are used to represent different writers.

Vuurpijl et al. [17] proposed a new method for writer identification that uses an "edge based" feature extraction to recognize writers. This method gives best result for large amount of text and considerably it gives higher dimension.

Cheung et al. [18] proposes new method for offline Text- independent writer identification which uses a mixture of global features and local features. They work on the "Gabor feature and LBP feature "extraction method. The two basic commitments are: to begin with, we appear in area 3 that meta-information and content characteristics on the web page containing the picture give a helpful appraisal of the likelihood that the picture is in class, and thereupon can be utilized to effectively rank pictures in the downloaded pool.

III. PROPOSED METHOD

A. The Framework of the Proposed Method

The proposed method comprises of three stages: training, enrollment, and identification, as appeared in Fig. 1. The proposed system consists of four stages. The stages are enrollment, training and identification.

Firstly, the handwriting sample image is segmented into word regions. After getting the word regions the scale invariant feature transform (SIFT) algorithm is applied to word regions. The SIFT is extract the key points, scales and orientations and key descriptors. The key descriptors generated after key points are computed. In all the stages the word segmentation is common.

In the training stage, firstly the SIFT features are extracted from word regions and the SIFT descriptors (SDs) are calculated. These descriptors are collected and generate the codebook. This codebook is used in enrollment and identification stages. In the enrollment stage, two features, called SIFT descriptors signature (SDS) and scale and orientation histogram (SOH), are extracted from SIFT descriptors and Scale and orientations of word regions of the enrolling handwriting image and stored for identification. In the identification stage, the SDS and SOH are extracted from the input handwriting images and respectively matched with the enrolled ones to get two matching distances, which are then fused to form the final matching distance for decision. As shown in Fig. 1, there are four main parts in the framework, i.e. word segmentation, codebook generation, feature extraction, and feature matching and fusion.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 90-94
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

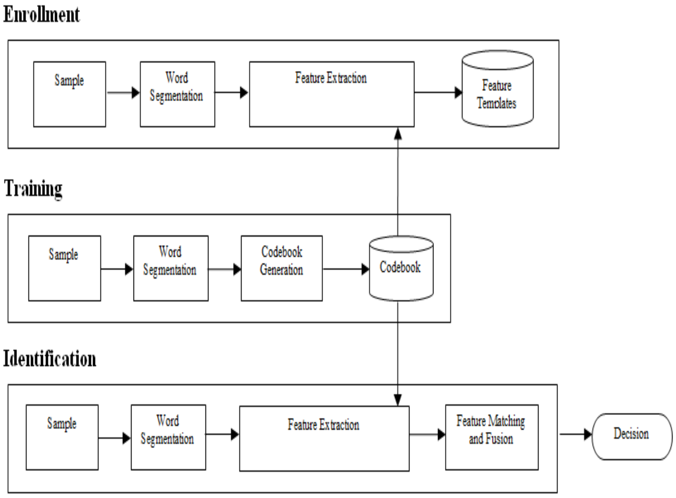


Fig.1.The framework of the proposed method

B. Word Segmentation

In early days the handwritten image is identified manually [19], which is exceptionally time consuming and tedious. To do handwriting image analysis segmentation is important. Recently many methods are introduced for handwriting identification in word level.

In word segmentation firstly the given image is divided into segments and extract the structural features of given image. Some authors are segmenting the full text lines, it does not support for skew handwriting images. To overcome this type of failures here we use a Laplacian of Gaussian (LoG) filtering.

Word segmentation has following steps:

1. Taking a one input image and convert it into binary image using Otsu's algorithm.
2. Getting every connected component (CCs) and calculate the average height.
3. Isotropic LoG filter is applied to binary image.
4. Binarizing the filtered image using Otsu's algorithm.
5. Allocating each associated words of binary image to the associated.
6. Combining the Semi word regions to get the word local points.
7. Collect all the connected components and finally get the segmented word.

C. SIFT

Scale invariant feature transform (SIFT), exhibited by Lowe [20] for particular distinctive scale invariant features extraction from images, has been widely and successfully applied in many fields [21].

This algorithm has following steps:

- Scale-space construction,
- Key point localization,
- Orientation assignment,
- Key point descriptor extraction.

After the word segmentation the SIFT algorithms is applied to word regions of input handwriting image. Here the above steps are carried out one by one, firstly the word regions of the segmented word takes as input.

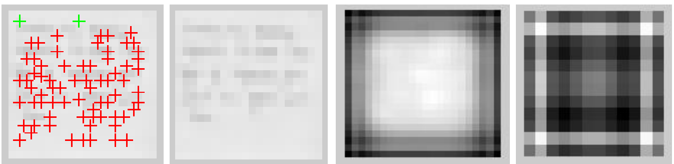
The Gaussian pyramid is applied to decompose the input handwriting and this pyramid of each part is called an octave. The levels are decomposed until the words are completed. In second and third step the local key points and scales are detected. In the fourth step the descriptors of key points are calculated.

D. Codebook generation

From a given image, many word regions (WRs) are acquired after word segmentation. For each WR, we apply the SIFT algorithm to recognize various key points, scales and orientations. Fig. 2 shows the key point localization, scale space description, key point descriptors and orientations.

We may get a large number of key points from different handwriting images expansive and differing measure of key focuses from various handwriting images.

It is difficult to maintain the key points for this reason clustering is done. This cluster consists of SDs and SOs of given hand writing images. We cluster the SDs of the key points extracted from the training samples into N categories and represent each category with its center, which is called a code. The greater part of the N codes frames a SD codebook with size N.



(a) (b) (c) (d)

Fig.2.(a) key points (b) key point descriptor (c) scale space (d) orientations.

In this work, the progressive Kohonen Self Organization method (SOM) clustering algorithm [22], which has been effectively utilized for codebook generation offline text-independent writer identification.

E. Feature Extraction

Once we get the SIFT keys and their descriptors we go for extracting the features. In feature extraction the signature of the descriptors are calculated and histograms are displayed in the form of graph. SIFT Descriptor Signature (SDS) Extraction and Scale and Orientation Histogram (SOH) Extraction are used to extract the features.

F. Feature matching and fusion

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 9, ISSN No. 2455-2143, Pages 90-94
Published Online July – August 2016 in IJEAST (http://www.ijeast.com)

Give two handwriting images as input, signify their SIFT descriptors signature and Taking two input images we compare the SDS and SOH of the handwriting images.

IV. EXPERIMENT AND RESULT

Using scale invariant feature transform (SIFT), the accuracy, recall, precision and time is calculated. Here six writer's samples are used and each writer consists of seven different samples.

Accuracy is used to calculate the overall correctness of the model and it is calculated as the ratio of the sum of correct classifications, as determined using the following equation.

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

These are the notations, which are used to calculate the Accuracy, Precision and recall.

Accuracy = (TP+TN) / (TP+TN+FP+FN)

The precision is the portion of the positive cases that were predicted correctly, and equation is

Precision = TP / (TP+FP)

The recall is the proportion of the correctly identified positive cases, and it is calculated using the equation.

Recall = TP / (TP+FN)

The table 1 shows the accuracy, recall, and precision and feature extraction time of six persons using seven samples of each person handwriting images. Comparing the values of SIFT and GLCM methods for most of the persons handwriting identified accurately by SIFT.

Table-1 Performance of the SIFT feature extraction

Writer	Accuracy (%)		Recall		Precision		Time(sec)	
	SIFT	GLCM	SIFT	GLCM	SIFT	GLCM	SIFT	GLCM
1	100	0.8571	1	1	1	0.8333	16.11	25.11
2	83.33	0.7142	0.83	0.8	1	0.8	13.10	23.73
3	100	85.71	1	0.8571	1	1	15.66	25.01
4	100	85.71	1	1	1	0.8333	18.07	29.46
5	100	100	1	1	1	1	16.11	27.54
6	100	100	1	1	1	1	16.95	27.61

V. CONCLUSION

This paper proposes a method for automatic offline text-independent writer identification based on SIFT. The SIFT method is a one of the structural based approach, in which key points are detected and description of are computed. Here SDS and SOH are extracted from the given handwriting images. Compare to the texture based approach like GLCM. The SIFT feature extraction method gives the better result. It is a word-level feature extraction of handwriting identification and also it is more suitable for identifying individual writers.

VI. REFERENCES

[1] E. Zois and V. Anastassopoulos, "Morphological waveform coding for writer identification," Pattern Recognit., vol. 33, no. 3, pp. 385–398, Mar. 2000.

[2] R. Jain and D. Doermann, "Offline writer identification using k-adjacent segments," in Proc. Int. Conf. Document Anal. Recognit., Beijing, China, 2011, pp. 769–773.

[3] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and Allographic features, " IEEE Trans. Pattern Anal Mach. Intell., vol. 29, no. 4, pp. 701–717, Apr. 2007.

[4] G. Louloudis, N. Stamatopoulos, and B. Gatos, "ICDAR 2011 writer identification contest," in Proc. Int. Conf. Document Anal. Recognition., Beijing, China, 2011, pp. 1475–1479.

[5] G. Louloudis, B. Gatos, and N. Stamatopoulos, "ICFHR 2012 competition on writer identification challenge 1: Latin/Greek documents," in Proc. Int. Conf. Front. Handwrite. Recognition., Bari, Italy, 2012, pp. 829–834.

[6] R. Plamondon and G. Lorette, "Automatic signature verification and writer identification—The state of the art," Pattern Recognit., vol. 22 no. 2, pp. 107–131, Feb. 1989.

[7] R. Hanusiak, L. Oliveira, E. Justino, and R. Sabourin, "Writer verification using texture-based features, " Int. J. Document Anal. Recognit.,vol. 15, no. 3, pp. 213–226, Sep. 2012.

[8] H. Said, T. Tan, and K. Baker, "Personal identification based on handwriting," Pattern Recognit., vol. 33, no. 1, pp. 149–160, Jan. 2000.

[9] Y. Zhu, T. Tan, and Y. Wang,"Biometric personal identification based on handwriting," in Proc. Int. Conf. Pattern Recognit., Barcelona, Spain,2000, pp. 797–800.

[10] B. Helli and M. Moghaddam, "A text-independent Persian writer identification based on feature relation graph (FRG)," Pattern Recognit., vol. 43, no. 6, pp. 2199–2209, Jun. 2010.

[11] Z. He, X. You, and Y. Tang, "Writer identification using global wavelet based features," Neurocomputing, vol.71, nos. 10–12, pp. 1832–1841,Jun. 2008.

[12] L. Du, X. You, H. Xu, Z. Gao, and Y. Tang, "Wavelet domain local binary pattern features for writer identification," in Proc. 20th Int. Conf. Pattern Recognit., Istanbul, Turkey, 2010, pp. 3691–3694.

[13] D. Bertolini, L. Oliveira, E. Justino, and R. Sabourin, "Texture-based descriptors for writer identification and verification," Expert Syst. Appl., vol. 40, no. 6, pp. 2069–2080, May 2013.

[14] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and Allographic features, " IEEE Trans. Pattern Anal Mach. Intell., vol. 29, no. 4, pp. 701–717, Apr. 2007.

[15] V. Papavassiliou, T. Stafylakis, V. Katsouros, and G. Carayannis, "Handwritten document image segmentation into text lines and words," Pattern Recognit., vol. 43, no. 1, pp. 369–377, Jan. 2010.

[16] C. Djeddi, I. Siddiqi, L. Souici-Meslati, and A. Ennaji, "Textindependent writer recognition using multi-script Hand written texts," Pattern Recognit. Lett., vol. 34, no. 10, pp. 1196–1202, Jul. 2013.

[17] M. Bulacu, L. Schomaker, and L. Vuurpijl, "Writer identification using edge-based directional features," in Proc. 7th Int. Conf. Document Anal. Recognition. (ICDAR '03), Piscataway, NJ, USA, 2003, pp. 937–941.

[18] Yiu-ming Cheung, Junping Dengion" Offline Text-Independent writer identification using a mixture of global and local features" 2011 Seventh International Conference on Computational Intelligence and Security.

[19] S. Srihari, S. Cha, H. Arora, and S. Lee, "Individuality of handwriting," J. Forens. Sci., vol. 47, no. 4, pp. 856–872, Jul. 2002.

[20] D. Lowe, "Distinctive image features from scale-invariant keypoints,"Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[21] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., New York, NY, USA, 2006, pp. 2161–2168.

[22] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.