ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

Identification of various writer's handwritten Text using Scale Invariant Feature Transform

Harshal Kolhe[1], Prashant Jadhav[2], Mahadev Kolape[3], Kunal Shah[4]

BE final year student, Computer Engineering, Sinhgad Institute of Technology, Lonavala, India[1,2,3]

Professor, Computer Engineering, Sinhgad Institute of Technology, Lonavala, India[4]

Abstract: Now a day's OCR helps us to recognize text from images but it is not operated with handwritten characters. So we proposed new methodology to recognize handwritten characters. In existing system LoG filter is used to segment the writer's handwriting image into word regions the SIFT descriptors (SDs) of WRs and the corresponding scales and orientations (SOs) are extracted which is part of SIFT have been utilized to recognize the features.

Keywords: Offline text independent writer identification, SIFT, word segmentation, SIFT descriptor signature, scale and orientation histogram.

I. INTRODUCTION

The handwriting text identification of writer is to determine the writer text among a number of known writers using hand writing images. In the human environment, it is very important to search the true writer of an unknown handwriting document. Writer identification has been important research topic in recognition field since several decades before. In our research, we have to find the global styles of various people's handwritings are obviously their handwriting images.

Author	Year	Feature Description	Feature Level	Domain	Accuracy (%)
Wang[9]	2003	Directional features	C	Offline	71
Pervouchine [14]	2006	Structural Features	C	Offline	72
Bulacu [3]	2007	Directional features	P	Offline	89
Neils [15]	2008	Allograph prototype matching	C	Offline	88
Chan [16]	2008	Discrete character prototype distribution	C	On-line	75
Tan [8]	2009	Continuous character prototype distribution	C	On-line	77

The approach existing for offline text independent writer identification can be roughly divided into two sub category.

1) Texture based approach
2) structure based approach

Texture approach take hand writing texts as a special texture image and extract the textual features for writer identification used a grey level co- occurrence matrix (GLCM)extract textual feature. From the handwriting images this extracted feature based on hidden markow tree(HMT) model wavelet domain for writer identification. The extracted wavelet based textual feature from hand writing images then it used Gabor and Xgabor and employed a feature relation graph(FRG) to represent extracted feature. It considered both local binary pattern (LBP) and local phase quantization (LPQ) as for texture descriptors of handwriting for writer verification and identification.

II. LITERATURE SURVEY

Table (C and P respectively means Character, Paragraph) Bulacu et al. [3] propose a texture-level approach using edge based probability distribution functions PDFs as features for text-independent writer identification task. Edge-hinge distribution introduced as a new feature. The key idea behind this feature is to consider two edg-fragments in the neighborhood of a pixel and compute the joint probability distribution of the orientations of the two fragments. Chan[16] take two pages of handwritten text as input and determine if they have been produced by the same writer. The features used to characterize a page of text include writing slant and skew, character height, stroke width, frequency of loops.

A. Existing System

In existing system SIFT have been utilized to recognize the features. Offline text-independent writer identification method based on scale invariant feature transform (SIFT) algorithm, is a accumulation of enrollment, training and identification stages. In all stages of proposed system, an isotropic LoG filter is used to segment the handwriting image into word regions (WRs).Then, the SIFT descriptors (SDs) of WRs and the corresponding scales and orientations (SOs) are extracted.

B. Proposed System

In existing system SIFT algorithm is used for extracting the features, but the problem with existing system is that accuracy getting by applying SIFT on various handwriting images for extracting the features is somewhat not efficient or unable to produce proper result. So, we proposed a system using same SIFT to improve accuracy for getting better result and is more efficient to user. We segment the project into three stages.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

The new proposed system having 3 stages i.e. Training, enrolment and identification. In Enrolment stage two features, are used i.e SD signature(SDS) and SO histogram (SOH), which are extracted from SDs and SOs of WRs of the enrolling handwriting image and stored for identification stage. In the training stage, an SD codebook is generated by clustering the SDs of training samples. In the identification stage, the SD signature and SO histogram are extracted from the input handwriting images and matched with the enrolled ones to get two matching distances, which are then fused to form the final matching distance for decision and result.

III. METHODOLOGY

In all these three stages handwriting images segmented into word regions(WRs). Then SIFT used to identify key point and get SIFT descriptors (SD) & corresponding scales and orientation(SO) from WRs. Both SD (SIFT Descriptor) and SO(SIFT Orientation) are very important information of handwriting to distinguish different writers. Therefore, in the following subsections, these SIFT information will be used to extract features of handwriting for writer identification. Both SD and SO are very important information of handwriting to distinguish different writers.

Therefore, in the following subsections, these SIFT information will be used to extract features of handwriting for writer identification. The SDs & SOs will be used in different ways in different stage. In the training stage SD's extracted from the training data set are used to generate code book for the use of enrolment and identification. In enrolment two feature called SD signature (SDs) & SO histogram(SOH) are used from SDs & SOs of WRs of the enrolling handwriting image & stored for identification. In identification stage SDs & SOH are extracted from the input handwriting image and respectively matched with the enrolled once to get two matching distance which are then fused to form final matching distance from decision.

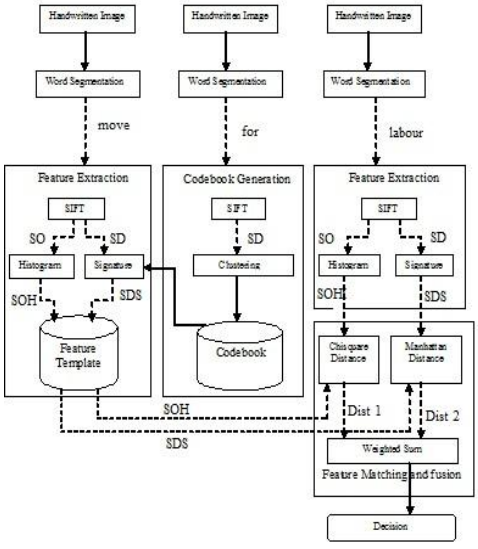

Fig 1: Modified Block Diagram of Proposed System.

A. *Word Segmentation*

The handwriting image I converting into word segmentation process can be simply following procedure.

1) Converting handwritten image (I) to binary image using OTSU's algorithm.
2) Getting all connected components (Ccs) in I_{bi} and then computing their average height h_{avg}
3) Filtering I_{bi} with an isotropic LoG filter to get the filtered image I_{fi}. The average height h_{avg} of all Ccs in I_b to decide the variance of the filter as $\sigma=2.5* h_{avg}$
4) Binarizing I_{fi} to get a binary image I_{fb} by using threshold obtained by OTSU's algorithm.
5) Assigning each connected component in I_{bi} to the nearest connected regions of I_{fb} to form semi word regions (SWR) which colored different.
6) Merging the SWR's to induce the word regions in line with the gap between the adjacent SWR's.
7) Splitting the overlapping Connected Components runs along multiple text lines from middle line of these boundary box.

B. *SIFT Algorithm*

SIFT Algorithm has 4 major stages of computation.

1) Scale space construction.
2) Key point localization
3) Orientation assignment
4) Key point descriptor extraction.

In the first stage the original images area decomposed into a Gaussian pyramid and each level of the pyramid called an octave. Which is further decomposed into several sub levels by convolving the initial image at the corresponding pyramid level with DoG filters with different variances. In second and Third stage many stable key point are detected and the location, scales and orientations of these key point are computed. In last step SIFT descriptor for each key point is generated.

In this work, we use SIFT to get the key points of handwriting, their SIFT descriptors (SDs), and the corresponding scales and orientations (SOs). The SDs are scale and rotation invariant and can reflect the structures of the image regions centered at the key points and the SOs can preserve the scale and orientation information of these structures. Both SD and SO are very important information of handwriting to distinguish different writers. Therefore, in the following subsections, these SIFT information will be used to extract features of handwriting for writer identification.

C. *Code book Generation*

For each word regions we used SIFT algorithm for detect no of key point and extract their descriptors scales and orientations. In code generation hierarchical Kohonen SOM clustering algorithm is used.

All of N codes form a SDs code book with size N and based on the code book we will compute a histogram with limited and fixed dimension as feature vector for writer identification.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015

D. Feature Extraction

Sometime the text in the identification hand writing document may be totally different with the text in the enrolled hand writing document in project. At that time the layout of the key point may be totally different in the different hand writing images even if same written by same person. So, for mathematical calculation not considering key point considering feature extraction and matching by recalculating frequency each SD & SO occurrence in the handwriting image.

IV. EXPECTED RESULTS

Our analysis work gift 3 section for author identification by victimisation improved feature Extraction technique in SIFT. Our planned system will provide additional correct user identification by obtaining completely different sample of same user will improve the result for author identification. Our system will increase the system's accuracy by applying SIFT in author identification purpose.

V. CONCLUSION

As per the proposed methodology we can identify various users with respect to their handwriting using SIFT algorithm which produce various features accurately. The proposed system computes the frequency of local structure features occurrences in a handwriting image and the local structures of these few special strokes make very little contribution to feature extraction. so we can expect more accuracy in writer identification using SIFT.

REFERENCES

[1] H. Said, T. Tan, and K. Baker, "Personal identification based on handwriting," Pattern Recognit., vol. 33, no. 1, pp. 149–160, Jan. 2000.
[2] L. Schomaker and M. Bulacu, "Automatic writer identification using connected-component contours and edge-based features of uppercase Western script," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 6,pp. 787–798, Jun. 2004.
[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] C. Djeddi, I. Siddiqi, L. Souici-Meslati, and A. Ennaji, "Text independent writer recognition using multi-script handwritten texts,"Pattern Recognit. Lett., vol. 34, no. 10, pp. 1196–1202, Jul. 2013.
[5] D. Lowe, "Distinctive image features from scale-invariant keypoints,"Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.
[6] T. Kohonen, "The self-organizing map," Proc. IEEE, vol. 78, no. 9,pp. 1464–1480, Sep. 1990.
[7] 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.
[8] 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.
[9] X.Wang,. Ding, H. Liu, "Writer identification using directional element features and linear transform," in Proceedings of the 7th International Conference Document Analysis and Recognition, 2003, pp. 942–945
[10] Y. Nakamura and M. Kidode, "Individuality analysis of online kanji handwriting," in International Conference on Document Analysis and Recognition, 2005.
[11] K. Steinke, "Recognition of writer by handwriting images," Pattern Recognition, vol. 14, no. 1-6, pp. 357–364, 1981.
[12] T.Daslin Pon Santhosh , P.KANNAN" A Literature survey for Object Recognition using Neural Networks in FPGA"(IJSRET) Volume 2 Issue 10 pp 677-682 January 2014.
[13] Khaled Mohammed bin Abd, siti zaiton mohd Hashim, Feature Extraction and Selection for Handwriting Identification,2003.
[14] Pervouchine, V., Leedham, G. (2006) ―Extraction and analysis of document examiner features from vector skeletons of grapheme 'th',‖ in Document Analysis Systems, pp. 196–207.
[15] Niels, R., Gootjen, F. Vuurpijl, L. (2008) "Writer Identification through Information Retrieval: The Allograph Weight Vector," in International Conference on Frontiers in Handwriting Recognition, pp. 481-486.
[16] Chan, S. K., Viard-Gaudin, C., Tay, Y. H. (2008)"Online writer identification using character prototypes distributions," inProceedings of SPIE - The International Society for Optical Engineering.

BIOGRAPHIES

Harshal Kolhe is student of BE in computer engineering from Sinhgad Institute of Technology, Lonavala, Pune affiliated to AICTE under Savitribai Phule Pune University and Completed Diploma in computer engineering in the year 2012 from Smt. Sharshchandrika Suresh Patil Institute Of Technology, Chopda, Jalgaon in MSBTE.

Prashant Jadhav is student of BE in computer engineering from Sinhgad Institute of Technology, Lonavala, Pune affiliated to AICTE under Savitribai Phule Pune University and Completed Diploma in computer engineering in the year 2012 from Government Polytechnic, Mumbai.

Mahadev Kolape is student of BE in computer engineering from Sinhgad Institute of Technology, Lonavala, Pune affiliated to AICTE under Savitribai Phule Pune University and Completed Diploma in computer engineering in the year 2012 from Gourihar Polytechnic , Limb, Satara in MSBTE.

Kunal Shah has received BE in Information Technology from, Sinhgad Institute of Technology, Lonavala, Pune affiliated to AICTE under Pune University, Maharastra, India in the year 2009 and ME in Computer Engineering from University Of Pune, India in the year 2014 also he is working as a asst. Prof. in department of Computer Engineering Sinhgad Institutes of Technology, Lonavala, Pune affiliated to AICTE under Savitribai Phule Pune University.