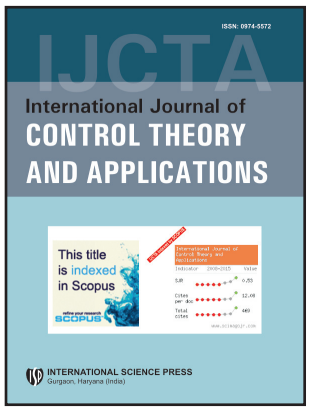

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 44 • 2016

Evaluation on the Usefulness of Texture and Binary Key-Point Features for Multi-Script Identification in Scene Images

Zaidah Ibrahim[a], Xiang Jian[b], Wenjing Chia[c], Muthukarupan Annamalai[d] and Dino Isa[e]

[a,d]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia. Email: [a]zaidah@tmsk.uitm.edu.my; [d]mk@tmsk.uitm.edu.my

[b,c]Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. Email: [b]Xiangjian.He@uts.edu.au; [c]Wenjiing.Jia@ut.edu.au

[e]Faculty of Engineering, University of Nottingham Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor. Email: dino.isa@nottingham.edu.my

Abstract: In this paper, we evaluate the usefulness of five texture and binary key-point features for multi-script identification in scene images. Automatic multi-script identification is essential for Optical Character Recognition (OCR) in a multi-lingual environment since most of the current OCRs are script dependent. Various researches have been developed for multi-script identification but mostly are for document images. Identification of scripts in scene images is more challenging since they contain varieties of font types, sizes and colours. The popular texture features used for text script identification are Speeded-up Robust Features (SURF), Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP). The popular features used for object recognition are binary key-point features such as Fast Retina Key-point (FREAK) and Binary Robust Invariant Scalable Key-points (BRISK). However, these five features have not been used to identify text script embedded in scene images. Thus, we attempt to investigate whether these features can be useful in scene images. The dataset consists of 500 images of text scripts in Malay, Chinese, Tamil and Jawi extracted from outdoor signage. The usefulness of features to classify the scripts in scene images are evaluated separately using multi-class Support Vector Machine (SVM) and K-nearest neighbour (KNN) classifiers. Results indicate that HOG has the highest script identification accuracy while BRISK extracts the features with the least processing time.

Keyword: BRISK, FREAK, HOG, LBP, SURF, multi-script identification, KNN and SVM.

1. INTRODUCTION

International travelling and communication have grown rapidly and travellers frequently encounter multi-scripts in scene images such as sign-boards and shop names. Script is a graphic form to represent a language. The objective of multi-script identification is to identify the script into one of several predefined languages. There are various challenges to perform multi-script identification from scene images. It is more challenging to recognize multi-scripts in scene images compared to document images since they are usually made up of various font types,

sizes and colours. The background of scene images is usually more complex and not just white background as in document images. This research caters for word-level script identification and this is another challenge since the length of the word for each script varies.

Nowadays, we can find documents, either machine printed or handwritten, and scene images such as sign-boards and shop names that consist of more than one scripst. This is a common situation especially in countries where English is not their first language. For instance, in Asian countries like Malaysia and Singapore, we can find documents and scene images that are written in English, Malay, Chinese, Tamil or Jawi. Figure 1 illustrates sample images of shop names in Malaysia that consists of multiple scripts. In Japan, we can find sign-boards written in English and Japanese, or English and Korean scripts in Korea. But if one does not understand any of the different scripts shown on the scene images such as sign-boards or street names, one may face problems in reaching one's destination or difficulty in having conversation with the native person of that country. Image understanding applications require text translation and recognition. But these can only be performed once the script has been identified and most Optical Character Recognition (OCR) only caters for a single script. Thus, an automatic multi-script identification is critical prior to character recognition so that it can be processed by the appropriate OCR and the recognition performance is optimized.

Figure 1: Sample scene images with multiple scripts.

Text script can be classified into three categories namely Latin, Ideography and Arabic [1]. Latin is written in languages like English or Malay. Ideography is written in languages like Chinese or Tamil while Jawi falls under the Arabic language. Latin and Ideography usually have similar height and pixel or stroke density while Arabic has lower height and density compared with the other two [1]. But these rules may be applied for text in document images since they are usually printed in a known font type and font size.

The performance of script identification depends on the features extracted from the word image. They should be informative and significant enough for the classifier to classify them into the correct type of script. One of the early researches in script identification for document images utilizing the optical density of the text image and classification is conducted by measuring the Euclidean Distance (ED) in Linear Discriminate Analysis (LDA) space from the centroids of the regions [2]. The results indicate that the Chinese and Tamil scripts have high level of density followed by Latin and Arabic. Shape characteristics is another feature that has been applied for printed and handwritten documents like contour and density [3] and curvature feature computed using bi-quadratic interpolation method [4]. Shape and density are sensitive to noise and since scene images are influenced by illumination changes, they not a suitable feature to be applied for scene images.

Texture feature is a popular feature for script identification because different scripts have visual appearance that leads to unique texture patterns compared to the background. Gabor filter [5], Co-occurrence Matrix of Oriented Gradients (Co-MOG) [6] and wavelet transform-based feature [7] have shown promising results. They are robust to font size and type and less prone to errors [1]. These are the feature characteristics that are suitable for scene images.

There are other texture features being applied for object recognition. For instance, Histogram of Oriented Gradient (HOG) for gait recognition [8] and hand gesture recognition [9], Local Binary Pattern (LBP) for face recognition [10] and Speeded-Up Robust Features (SURF) for vehicle recognition [11]. Besides that, there are binary key-point features for object recognition being produced that are Fast Retina Key-point (FREAK) [12] and Binary Robust Invariant Scalable Key-points (BRISK) [13]. Since these features illustrate good object recognition results, this research attempts to evaluate the usefulness of these features for script identification in scene images.

Two classifiers are being used for the evaluation process that is multi-class Support Vector Machine classifier (SVM) and K-Nearest Neighbor (KNN). A multi-script dataset is constructed that consists of 500 manually cropped words from the captured images of signboards, shop names and street names in Malaysia.

This paper is organized as follows. Section II explains the related work while section III discusses the methodology and experimental results of the comparative study. Section IV concludes this research and list future work.

2. RELATED WORK

A popular feature in computer vision is texture where it is powerful in various image processing applications. Texture is a structure composed of ordered similar patterns [15]. Among the popular texture features are HOG, LBP, SURF while popular binary key-point features are BRISK and FREAK.

HOG characterizes the script appearance by the distribution of local intensity gradients that are computed for each cell or regions divided in the image. Each cell's pixel contributes weighted gradient to its corresponding angular bin. Groups of adjacent cells are known as blocks. Normalized group of histograms represents the block histograms that are the result of the descriptor [16]. LBP labels the pixels of an image by thresholding the neighbourhood of each pixel and considers the result as a binary number [17]. SURF is a local invariant fast feature point detector and descriptor. It utilizes integral images and scale space construction to extract keypoints efficiently. Then, it assigns an orientation within a circular region around the keypoints. After that, a squared area is aligned to the identified orientation and the SURF descriptor is extracted using Haar wavelet. As a result, a 64-dimensional vector is produced [18].

BRISK calculates the weighted Gaussian average over a chosen pattern of points near the keypoint. It compares the values of specific pairs of Gaussian windows, giving a 1 or 0 output based on which window in the pair is bigger [19]. FREAK simulates human vision process by using a retinal sampling grid where higher density points are located at the centre of the sampling grid. A FREAK descriptor is generated by thresholding differences of comparable Gaussian kernels. A 64-dimensional vector is generated [20].

Support vector machine (SVM) seems to be a popular classifier where it has shown its good performance for script recognition [6], [13] and thus being used in this research. SVMs are supervised learning models with associated learning algorithms that analyse data and recognize patterns. Given a set of training examples, each belongs to one of two categories, making it a non-probabilistic binary classifier. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Support vector machine (SVM) seems to be a popular classifier where it has shown good performance for script recognition [14], [21]. SVMs are supervised learning models with associated learning algorithms that analyse data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each belongs to one of two categories, making it a non-probabilistic binary linear classifier. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. KNN is a non-parametric approach used for classification where the input contains the k closest training examples in the feature space. The output is a class membership where it is classified by a majority vote of its neighbours [22].

3. EXPERIMENTAL RESULTS

The implementation has been conducted in Matlab. Our dataset consists of Malay, Chinese, Tamil and Jawi scripts manually cropped from scene images like shop names, street names and signboards at word level. The accuracy of the multi-script identifier performance is measured by dividing the number of word images whose

scripts have been correctly classified with the overall images. 125 word text images from each of the four scripts have been manually cropped from scene images of shop names, street names and signboards which total up to 500 multi-script texts. Eighty percent of this multi-script text has been used for training and the remainder is being used for testing. Figure 2 illustrates some sample images of the multi-script words.

Corners are keypoints that have high curvature and exist at the junction of different brightness blocks of images. Detecting corners can reduce processing time without losing image data information. Since almost all scripts consist of corner keypoints, corner detection is also integrated in BRISK and FREAK. Three popular corner detectors are Harris [23], Eigen and Features from Accelerated Segment Test (FAST) [24]. Harris corner considers the differential of the autocorrelation with respect to the edge direction. A point is classified as a corner in FAST if a sufficiently large set of pixels can be detected on a circle of fixed radius around the point such that these pixels are all significantly brighter than the central point.

These experiments have been conducted on an Intel CORE i5 with 4GB RAM. Table 1 illustrates the result of the feature extraction process for each feature per image. The time is measured in seconds. By looking at Table 1, we can see that shortest time for features to be extracted are by BRISK and FAST corner detector since BRISK renders lower memory load. Table 2 shows the script recognition rate classified by KNN while the results produced by SVM classifier are shown in Table 3. By comparing the results in Table 2 and Table 3, we can see that HOG has the highest script recognition rate for both KNN and SVM. This is because as a gradient-based feature, HOG really captures the appearance information of the characters. This is produced by dividing the image into small connected regions and for the pixels within each region; a histogram of gradient directions is extracted. The output of HOG is the concatenation of these histograms. Thus, it is invariant to geometric and photometric transformations since changes appear in larger spatial regions.

Results in Table 2 and Table 3 also show that Jawi script can be easily recognized compared to the other scripts because there is not many variations in terms of the shape of the characters. Please refer to Figure 2 for an illustration of some samples of Jawi scripts. The other scripts have poor results due to the variations of font style.

Figure 2: Sample images of multi-script words

Table 1
Processing time for feature extraction for one image (sec)

	HOG	SURF	LBP	FREAK (HARRIS corner)	FREAK (EIGEN corner)	FREAK (FAST corner)	BRISK (HARRIS corner)	BRISK (EIGEN corner)	BRISK (FAST corner)
Chinese	0.01200	0.00091	0.00036	0.03855	0.03573	0.02927	0.02409	0.00100	0.00091
Jawi	0.00782	0.00100	0.00027	0.03236	0.03227	0.03227	0.00073	0.00082	0.00073
Malay	0.01336	0.02145	0.00818	0.03291	0.03200	0.03318	0.00055	0.00082	0.00064
Tamil	0.00964	0.00091	0.00027	0.03291	0.03236	0.03264	0.00073	0.00082	0.00073
Average	0.01070	0.00607	0.00227	0.03418	0.03309	0.03184	0.00652	0.00086	**0.00075**

Table 2
Script recognition rate by KNN

	HOG	SURF	LBP	FREAK (HARRIS corner)	FREAK (EIGEN corner)	FREAK (FAST corner)	BRISK (HARRIS corner)	BRISK (EIGEN corner)	BRISK (FAST corner)	Average
Chinese	90.9	63.6	72.7	81.8	90.9	81.8	81.8	54.5	72.2	76.68889
Jawi	100	90.9	100	100	90.9	81.8	100	100	100	95.95556
Malay	100	54.5	90.9	27.3	63.6	36.4	36.4	45.5	45.5	55.56667
Tamil	90.9	63.6	72.7	54.5	72.7	54.5	81.8	90.9	72.7	72.7
Average	95.5	68.2	84.1	65.9	79.5	63.6	75	72.7	72.6	

Table 3
Script recognition rate by SVM

	HOG	SURF	LBP	FREAK (HARRIS corner)	FREAK (EIGEN corner)	FREAK (FAST corner)	BRISK (HARRIS corner)	BRISK (EIGEN corner)	BRISK (FAST corner)	Average
Chinese	81.8	45.5	.3 (65.	72.7	72.7	72.7	90.9	63.6	63.6	70.4375
Jawi	100	72.7	90.9	90.9	90.9	100	90.9	90.9	90.9	90.9
Malay	100	54.5	81.8	72.7	72.7	81.8	36.4	45.5	45.5	65.65556
Tamil	72.7	27.3	63.6	90.9	90.9	72.7	63.6	81.8	81.8	71.7
Average	88.6	50	65.9	81.8	81.8	81.8	70.5	70.5	70.5	

4. CONCLUSION

An automatic multi-script identification is essential prior to character recognition processed by an OCR. Multi-script identification for document images is less challenging compared to scene images since the font can come in various font types, sizes and the inter-character spacing is also varies. For applications where memory and speed are the constraints, one needs to decide which factor is more important. Experiments indicate that HOG identifies the scripts better but at a higher speed while BRISK produces the result faster but with less accuracy. Future work involves the integration of multiple features for better recognition rate and processing time.

Acknowledgment

First author wishes to thank Universiti Teknologi MARA, Malaysia and University of Technology, Sydney, Australia, for sponsoring this research.

REFERENCES

[1] Ghosh, D.; Dube, T. and Shivaprasad, A.P., "Script Recognition – A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.

[2] Spitz, A. L., "Determination of the Script and Language Content of Document Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19. No. 3, March 1997.

[3] Kessentini, Y.; Paquet, T. and Hamadou, A.B., "A Multi-Lingual Recognition System for Arabic and Latin Handwriting", 10th ICDAR 2009.

[4] Chanda, S.; Pal, U. and Franke, K. "Font Identification – In Context of an Indic Script", 21st International Conference on Pattern Recognition (ICPR2012).

[5] Pati, P. B. And Ramakrishnan, A. G., "Word Level Multi-script Identification", Pattern Recognition Letters (2008).

[6] Saidani, A.; Kacem, A. and Belaid, A., "Co-occurrence Matrix of Oriented Gradients for Word Script and Nature Identification", 13th International Conference on Document Analysis and Recognition (ICDAR2015).

[7] Busch, A.; Boles, W.W. and Sridharan, S., "Texture for Script Identification", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 11, November 2005.

[8] Liu, Y.; Zhang, J.; Wang, Chen and Wang, L., "Multiple HOG templates for gait recognition", 21st International Conference on Pattern Recognition (ICPR2012).

[9] Prasuhn, L.; Oyamada, Y.; Mochizuki, Y. and Ishikawa, H., "A HOG-based Hand Gesture Recognition System on a Mobile Device", IEEE International Conference on Image Processing, (ICIP2014).

[10] Nikan, S. and Ahmadi, M., "Human Face Recognition under Occlusion using LBP and Entropy Weighted Voting", 21st International Conference on Pattern Recognition (ICPR2012).

[11] Hsieh, J.; Chen, L. and Chen, D., "symmetrical SURF and Its Applications to Vehicle Detection and Vehicle Make and Model Recognition", IEEE Transactions on Intelligent Transportation Systems, Vol. 15, No. 1, February 2014.

[12] Malik, M. I.; Ahmed, A.; Liwicki, M. And Dengel, A., "FREAK for Real Time Forensic Signature Verification", 12th International Conference on Document Analysis and Recognition (ICDAR2013).

[13] Ege, Y.; Nazlibilek, S.; Kakilli, A.; Citak, H.; Kalender, O.; Karacor, D.; Erturk, K. L. And Sengul, G., "A Study on the Performance of Magnetic Material Identification System by SIFT-BRISK and Neural Network Models", IEEE Transactions on Magnetics, Vol. 52, No. 8, August 2015.

[14] Ferrer, M.A., Morales, A. and Pal, U.; "LBP Based Line-wise Script Identification", 12th International Conference on Document Analysis and Recognition ICDAR 2013.

[15] Goot, L.V.; Dewaele, P. and Oosterlinck, A.; "Survey : Texture Analysis", Computer Vision, Graphics and Image Processing 29, 1985.

[16] Dalal, N. and Triggs, B., "Histograms of Oriented Gradients for Human Detection", IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[17] DC. He and L. Wang, "Texture Unit, Texture Spectrum, And Texture Analysis", IEEE Transactions on Geoscience and Remote Sensing, Vol. 28, 1990.

[18] Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., "SURF: Speeded-Up Robust Features", Computer Vision and Image Understanding, Vol. 110(3), pp. 346-359, 2008.

[19] Leutenegger, S.; Chli, M. and Siegwart, R.Y., "BRISK: Binary Robust Invariant Scalable Keypoints", International Conference on Computer Vision (ICCV2011).

[20] Alahi, A., Ortiz, R. and Vandergheynst, P., "FREAK: Fast Retina Key-Point", IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[21] Sharma, N.; Chanda, S.; Pal, U. and Blumenstein, M., "A study on Word-level multi-script identification from video frames", IEEE International Joint Conference on Neural Network 2014.

[22] Altman, N. S., "An introduction to kernel and nearest-neighbour nonparametric regression", The American Statistician. **46** (3), 1992.

[23] C. Harris and M. Stephens, "A Combined Corner and Edge Detector" (PDF), Proceedings of the 4th Alvey Vision Conference, 1988.

[24] Rosten, E.; Porter, R. And Drummond, T., "Faster and Better: A Machine Learning Approach to Corner Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 1, 2010.