Journal of the Korea Academia-Industrial cooperation Society
Vol. 19, No. 1 pp. 609-614, 2018

https://doi.org/10.5762/KAIS.2018.19.1.609
ISSN 1975-4701 / eISSN 2288-4688

Hierarchical Text Extraction and Localization on Images

Byoung-Min Jun[1*], Woogyoung Jun[2]

[1]**Department of Computer Engineering, Chungbuk National University**

[2]**Department of Computer Science, Yonsei University**

이미지로부터 계층적 문자열 추출에 관한 연구

전병민[1*], 전우경[2]

[1]충북대학교 컴퓨터공학과

[2]연세대학교 컴퓨터과학과

Abstract This study was conducted to investigate the effects of turmeric powder on jeung-pyun. Turmeric jeung-pyun containing 0%, 0.5%, 1%, 1.5%, and 2% turmeric powder was prepared and the moisture, pH, sugar, color, texture, DPPH and sensory properties of the samples were measured. Moisture contents of jeung-pyun were 51.26~51.99% and there were significant differences among the samples($p<0.001$). The L-values were significantly decreased with increasing turmeric powder content. The b-value was low in the control and there were significant differences among the samples($p<0.05$). Texture profile analysis showed that there were no significant differences among the groups in hardness, adhesiveness, springiness, cohesiveness, gumminess, and chewiness. The hardness was the lowest in the control group and increased with increasing turmeric powder content. The antioxidant activities as measured by DPPH increased with increasing turmeric powder content ($p<0.001$). In the sensory evaluation, 1% addition of turmeric powder showed the highest preference in terms of color, taste, flavor, texture and overall preference($p<0.001$). As determined by this study, the addition of 1% turmeric powder was the most favorable method for making use of turmeric powder in the production of jueng-pyun.

요 약 인터넷 기술의 급격한 성장으로 우리들은 언제 어디에서나 다양한 장치를 이용하여 온라인에 접속할 수 있으며, 실시간, 대용량의 영상 및 사진들이 인터넷상에 올려지고 있다. 이러한 영상들의 대부분은 영상에 관련된, 영상을 인식할 수 있는 간단한 주석을 갖는다. 그럼에도 아직도 주석이 없는 단일 영상이나 잘못된 주석이나 태그 정보 때문에 우리가 원하는 영상을 찾는데 문제점이 있어 이러한 문제해결을 위해서는 영상의 올바른 정보를 태깅하는 것이 필수적이다. 대부분의 태그는 문서나 주석의 형태를 가지므로 주석이나 문서의 정보가 올바르지 않으면 원하는 영상을 찾는데 많은 어려움이 따른다. 그리하여 더 나은 영상 탐색 결과와 올바른 영상 주석을 위해서 작가에 의한 주석뿐만 아니라 올바른 영상분석 또한 아주 중요하다. 영상 특징을 추출하는 것은 신뢰성 있는 영상 주석을 위해 필수 불가결한 요소이다. 따라서 본 논문에서는 다양한 불특정 영상으로 부터 계층적 텍스트 추출 방법을 사용하여 신뢰성 있는 영상 주석을 얻는다. 다양한 영상으로 부터 영상이나 사진 속에 포함된 텍스트 정보를 추출하는 방법을 제안하였으며, 실험결과 제안한 텍스트 추출기법이 대부분의 영상으로부터 정확하게 텍스트 특징을 추출하는 결과를 보여주었고, 성능 평가 결과 최소 0.04부터 최대 0.52의 높은 평가결과를 보여주었다. 또한 정확도 측면에서도 다른 기법들 보다 최소 18.1%부터 최대 37.9%의 높은 정확도를 보여주었다.

Keywords : Image Processing, Pattern Recognition, Text Detection, Text Extraction, Text Localization

This work was supported by the intramural research grant of Chungbuk National University in 2015.

*Corresponding Author : Byoung-Min Jun(Chungbuk National Univ.)
Tel: +82-43-261-2453 email: bmjun@cbnu.ac.kr

Received November 10, 2017 Revised December 6, 2017
Accepted January 5, 2018 Published January 31, 2018

1. Introduction

Extracting text features on images are very important for image annotation and image retrieval as well. Most of images are explained with supporting document, comments, and tag information. Nevertheless, still tremendous number of images does not contain textual documents, tags, or comments. Even if the image has supporting related information, some images are still hard to annotate because of incorrect information. These makes hard to determine what does image means. In this case, we can annotate image meanings using image features. Image contains many useful features we can expect what image means. For example we can extract text information features from images. Nowadays, text extracting methods are applied on various real world areas such as vehicle license plate recognition and speed limit road sign recognition as shown in Figure 1. Most of real world text extracting has regular format. But in this case, we can't predict where does text feature locates on images. Find out where does text feature locates and extracting text features are important for tagging images and better image retrieve results. Therefore, we present hierarchical text extracting and localization method to extract text features from images.

Fig. 1. Example of text detection: License plate recognition and speed limit road sign recognition

In Section 2, we discuss previous works and present our three step hierarchical text extracting method: word level text recognition-character level text recognition-text localization in Section 3. We present our successful experimental result in Section 4. Finally we present conclusion of hierarchical text extraction and localization method and discuss feature works in Section 5.

2. Related Works

These days, various researches focus on text recognition and extraction. We can find out many other different approaches. Basically, there are some researches on recognition of words or letters in image that only contains text feature, [1,2]. We can find different approach called end-to-end text recognition in [3,4]. However, most important problem is how to separate or determine the region that contains text features. Most of researches for text recognition can be classified into two categories which is learning based method and rule based method. Rule based methods extract connected regions and then design the rules to separate text from other regions. Using connected component analysis, [5,6]localized the text in images. Stroke width filter based text extraction method is proposed in [7]. [8]proposed adaptive threshold and the rules of text recognition and won the first prize in competition. We can find more rule based method such as [9]using coarse-to-fine video text detection, localization, and extraction for multilingual text localization. [10]proposed heuristic rule based approach and [11]proposed corner based approach to localize text regions automatically. Using selective metric based clustering, [12]extracted color texts in images. Novel coarse-to-fine algorithm based on multi-scale wavelet features is proposed by [13] and [14] proposed a stroke width transform. However, there are limitations of rule based approach. Rule based approach is not robust and hard to clear background regions from the text.

Therefore, many researchers researched learning based method which is robust with training data. [15]made text statistic to select the discriminative features. [16] proposed framework based on SVM to perform accurate text localization and designed stroke filter combined with a SVM classifier to extract text regions on [16]. Analysis on text features in natural images using SVM classifier and locating the text region is proposed on [17],[18] proposed a systematic approach using HOG classifier to computer the

confident value of a region and then employed the pictorial structure to constrain the text spatial layout.

Fig. 2. Framework of our hierarchical text extraction method

3. Text Recognition and Localization

Unlike these related works, our approach is basically based on learning based method and Figure 2 shows the hierarchical framework of our approach. We combine word level features and character level features for text localization. In first step, called word level text recognition, we use stroke width transform for image filtering. In second step, called character level text recognition, we select the discriminative features and compute the confident value of the candidate region using random forest. Finally, we implement conditional random field to separate text from the background region, text localization.

3.1 Word Level Text Recognition

In a word level text recognition, we apply stroke width transformation since most of text have similar stoke width. There are no difference on original image size and stroke width transformation image and each pixel value as well. Fist, an image is processed by canny edge operator and gradient orientations. If pixel p is located at the edge of stroke, gradient orientation d_p is perpendicular to the stroke orientation. The radial $r = p + n \times d_p$ will cross at the other edge of the stroke. If the cross point at the other edge of the stroke was defined as q and w is the maximum value of the stroke width, the gradient orientations of p and q are inverse. Radial r will be removed if p does not find the corresponding point q at the other edge of the stroke. Figure 3 shows the stroke width transformation results.

Fig. 3. Original image on up-left, edge detected image on up-right, gradient orientation image on bottom-left, and stroke width transformation image on bottom-right

3.2 Character Level Text Recognition and Text Localization

Normalized gray intensity, color cues, constant gradient variance, and histogram of gradient features are typically used in text localization. We use histogram of oriented gradient method to describe the text. Histogram of oriented gradient method is a three dimensional histogram. We normalize a character image to 48×48 and the normalized image is tiled with the overlapping blocks of 5×5. We slide cell every pixel and for each cell, the magnitudes of image gradients weight orientation histogram with 9 bins. Histogram in each cell is being normalized. Finally, character image is represented by a feature vector with 48×48×9 dimensions.

Random forest classifier is a popular classification method. It consist many decision trees and the label of query image is determined by majority of the votes. Random forest is made up of several decision trees. The data in training set T is denoted as $x_1, x_2, \ldots, x_n$, $y_i \in T$ where x_i is the i^{th} dimension of feature vector and y_i is the label. Each tree is constructed by recursively splitting T into two subsets T_l and T_r. Feature x_i can be the best split of the T and it can result in the largest expected information gain of the node categories.

In the training step, we collected letter images from

ICDAR training data sets. We randomly sample the background as the negative samples as well. Our samples of training sets are shown in Figure 4.

We select discriminative features and compute the reliable values of the candidate regions using random forest. We implement sliding window method on the candidate regions on stroke width transformation image and then compute the confident weight based on the trained random forest classifier. Finally, we obtain a confident map where each pixel value is equal to its confident value.

In text localization step, we normalize the discriminative information of texts within text spatial layout based on conditional random field. Spatial layout in a word is characteristic for text recognition.

Fig. 4. Samples of ICDAR training data sets

4. Experiments and Results

We have done the experiment with ICDAR dataset which contains various neutral scene images such as landscapes, road signs, posters, and more. Figure 5 shows the visual results of text extraction using our hierarchical text extraction and localization method with ICDAR datasets in red rectangle. It localized exactly where text exists. It means that our approach is reliable and accurate.

Our approach efficiency is measured by evaluate functions such as recall, precision, and F-score and compared with 4 different methods introduced in Section 2. Performance comparison by recall, precision, and F-score is shown in Table 1 and accuracy comparison result is shown in Table 2. As shown in Table 1, our approach showed best performance in all evaluate functions and also shown in Table 2, our approach is 89.5% accurate no worst then others.

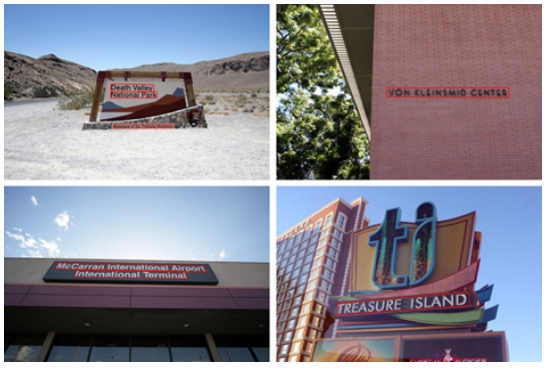

Fig. 5. Examples of text extraction and localization results

Table 1. Performance comparison

Method	Precision	Recall	F-Score
Our Method	**0.71**	**0.68**	**0.69**
Hinnerk	0.62	0.63	0.65
Ashida	0.55	0.46	0.50
Wolf	0.30	0.44	0.35
Todoran	0.19	0.18	0.18

Table 2. Accuracy comparison

Method	Accuracy(%)
Our Method	**89.5**
Hinnerk	71.4
Ashida	69.8
Wolf	53.1
Todoran	51.6

5. Conclusions

In this paper, we presented a hierarchical text extraction and localization method on neutral scene images. Our research showed that we can recognize and extract text region on images and locate the text region to ease annotate images as tags. Furthermore, it can be used for automatic image annotation for reliable image retrieval. We compared our text extract and localization method in terms of precision, recall, and F-scores and compared with 4 different approaches. Our research demonstrates that our method shows best precision, recall, and F-score. And the results also showed that our approach is most reliable and accurate. Since still there are images that we could not extract text features yet such as skewed text features or interruption over text features as shown in Figure 6, further research that can improve such problems must be proposed.

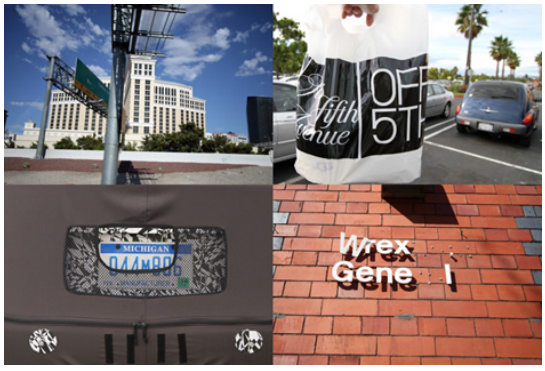

Fig. 6. Examples of unrecognized text extraction

References

[1] Wang, K., Belongie, S., "Word Spotting in the Wild", *ECCV 2010, Part I. LNCS*, vol. 6311, pp.591-604, 2010.

[2] Mishra, A., Alahari, K., Jawahar, C.V., "Top-Down and Bottom-Up Cues for Scene Text Recognition", *CVPR*, 2010.

[3] Wang, K., Babenko, B., Belongie, S., "End-to-End Scene Text Recognition", *2011 IEEE International Conference on Computer Vision*, pp. 1457-1464, 2011.
DOI: https://doi.org/10.1109/ICCV.2011.6126402

[4] Neumann, L., Matas, J., "A Method for Text Localization and Recognition in Real-world Images", *ACCV 2010, Part III. LNCS*, vol. 6494, pp. 770-783, Springer, Heidelberg, 2011.
DOI: https://doi.org/10.1007/978-3-642-19318-7_60

[5] Lienhart, R., Effelsberg, W., "Automatic text segmentation and text recognition for video indexing", *TR-98-009, University of Mannheim*, 1998.

[6] Jung, K., Kim, J., "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(12), pp. 1631-1639, 2003.
DOI: https://doi.org/10.1109/TPAMI.2003.1251157

[7] Liu, Q., Jung, C., Moon, Y., "Text Segmentation based on Stroke Filter", *Proceedings of International Conference on Multimedia*, pp. 129-132, 2006.
DOI: https://doi.org/10.1145/1180639.1180677

[8] Lucas, S.M., "ICDAR2005 text locating competition results", *Proceedings of the 8th International Conference on Document Analysis and Recognition*, pp. 80-84, 2005.

[10] Lyu, M.R., Song, J., Cai, M., "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction", *IEEE Transaction on Circuits and Systems for Video Technology*, pp. 243-255, 2005.

[9] Shivakumara, P., Huang, W., Phan, T.Q., "Accurate Video Text Detection through Classificaion of Low and High Contrast Images", *Pattern Recognition* 43, pp. 2165-2185, 2010.
DOI: https://doi.org/10.1016/j.patcog.2010.01.009

[10] Hua, X.-S., Chen, X.-R., Whnyin, L., Zhang, H.-J., "Automatic Location of Text in Video Frames", *Proceedings of the 2001 ACM Workshops on Multimedia*, 2001.
DOI: https://doi.org/10.1145/500933.500941

[12] Thillou, C.M., Gosselin, B., "Color Text Extraction with Selective Metric-based Clustering", *Computer Vision and Image Understanding 107*, pp. 97-107, 2007.
DOI: https://doi.org/10.1016/j.cviu.2006.11.010

[13] Ye, Q., Huang, Q., Gao, W., Zhao, D., "Fast and Robust Text Detection in Images and Video Frames", I*mage and Vision Computing 23(6)*, pp, 565-576, 2005.

[14] Epshtein, B., Ofek, E., Wexler, Y., "Detecting Text in Natural Scenes with Stroke width Transform", *2010 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2963-2970, 2010.

[15] Chen, X., Yuille, A.L., "A Time Efficient Cascade for Real-Time Object Detection", Proceedings of the CVAVI 2005, *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2005.
DOI: https://doi.org/10.1109/CVPR.2010.5540041

[16] Jung, C., Liu, Q., Kim, J., "Accurate Text Localization in images based on SVM output Scores", *Images and Vision Computing 27*, pp. 1295-1301, 2009.
DOI: https://doi.org/10.1016/j.imavis.2008.11.012

[17] Jung, K., Kim, J., "Texture-based Approach for Text Detection in Images using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm", *IEEE Transactions on Pattern Analysis and Machine*

Intelligence 25(12), pp. 1631-1639, 2003.
DOI: https://doi.org/10.1109/TPAMI.2003.1251157

[18] Wang, K., Babenko, B., Belongje, S., "End-to-End Scene Text Recognition", 2011 *IEEE International Conference on Computer Vision*, pp. 1457-1464, 2011.
DOI: https://doi.org/10.1109/ICCV.2011.6126402

Byoung-Min Jun [Regular member]

- Feb. 1978: Yonsei Univ., MS
- Aug. 1988: Yonsei Univ., PhD
- Feb. 1986 ~ current: Chungbuk National University. Dept. of Computer Engineering, Professor

<Research Interests>
Computer Vision, Image Processing, Pattern Recognition

Woogyoung Jun [Regular member]

- Feb. 2008: Yonsei Univ., MS
- Feb. 2015: Yonsei Univ., PhD

<Research Interests>
Artificial Intelligence, Pattern Recognition, Image Processing, Data Mining