March - April 2020
ISSN: 0193-4120 Page No. 4361 - 4368

Accurate Detection of Text Areas Using Fused Characteristics

Seok-Woo Jang[1], Sang-Hong Lee[*2],
[1]Assistant Professor, Department of Software, Anyang University, 708-113, Anyang 5-Dong, Manan-Gu, Anyang 14028, South Korea
[*2]Assistant Professor, Department of Computer Engineering, Anyang University, 708-113, Anyang 5-Dong, Manan-Gu, Anyang 14028, South Korea,
swjang7285@gmail.com[1], shleedosa@anyang.ac.kr[*2]

Article Info
Volume 83
Page Number: 4361 - 4368
Publication Issue:
March - April 2020

Abstract

Establishment and focus: Texts provide important information to convey the meaning of an image. Therefore, there is a need for accurately segmenting characters included in images as a prerequisite of character recognition.

System: This study introduces an approach of extracting text regions included in stereoscopic images based on texture and depth features. The method first segments candidate text areas using texture features. Then, after character region localization is performed, the background is separated from the localized character string. Finally, the depth feature is utilized to confirm whether the obtained text areas contain only the text regions and not the non-text regions. In the testing of this study, the introduced algorithm detected the character regions in input color images more accurately than the existing algorithm. To compare the performance of the introduced character region acquisition method, we used a correctness metric expressed as a percentage that shows the ratio of exactly localized strings that do not contain non-character areas to the total localized strings. For performance comparison, we also implemented the conventional neural network-based text detection method. In general, when there is an artificially inserted background, the background area is simple, so that the binarization of the text area is performed accurately. However, in the case of texts without an artificial background, it is difficult to accurately binarize the text and the background in the existing method, but the proposed method can obtain relatively good results by using texture and depth information simultaneously. The proposed method improves the accuracy of character area verification by greatly reducing the extraction of non-character regions with the help of three-dimensional depth information. The suggested text region extraction approach is expected to be very useful in computer vision related fields such as movie caption recognition, character recognition, license plate extraction, and so on.

Article History
Article Received: *24 July 2019*
Revised: *12 September 2019*
Accepted: *15 February 2020*
Publication: *26 March 2020*

Keywords: *Fused Characteristics, Color image, Background Area, Verification Stage, Text Area.*

1. Introduction

In recent years, various multimedia materials have been exponentially spreading due to the very fast Internet, the emergence of various authoring tools, and the development of mass storage devices. In general, multimedia materials are mixed contents that combine various media such as sound, video, and text. In particular, the character information such as the caption includes information that is very important for grasping the content or meaning of the image. Therefore, there is a continuous need for a study on accurately segmenting characters included in several moving images as a preliminary step of character recognition for indexing and searching multimedia materials [1-5].

In the literature, various research methods for segmenting character regions have been introduced. In the frequency-based method [6], after dividing the image into block units, the character was extracted using the feature that the high frequency elements horizontally and vertically are very large in character areas. The method proposed in the study [7] introduced a method of acquiring a text area by leaving the edge of the text area strongly and deleting the remaining edges. In the study [8-9], the saturation value is used and a color transfer map between the text and the background is generated to detect the character region. Additionally, other approaches of extracting the character area continue to be introduced [10].

Many techniques for segmenting characters have been studied, but there are still problems to be solved. In particular, many existing methods attempt to extract text areas mainly from two dimensional images. Recently, three-dimensional stereoscopic images have been widely used, but studies on extracting characters existing in three-dimensional images using three-dimensional features are currently hardly found.

Therefore, in this study, we introduce an approach of robustly extracting text areas included in color images based on texture and depth features by receiving three-dimensional stereoscopic images. Fig. 1 shows the flow chart of the character region acquisition strategy introduced in this study.

As can be seen in Fig. 1, the introduced algorithm accepts three-dimensional images taken in three dimensions and segments candidate character areas that are expected to have characters from the images using texture features. Then, after performing character region localization extracting only a portion forming the character string from the candidate character region, the background is separated from the localized character string and only the characters are extracted. Finally, the distance feature is utilized to confirm whether the obtained character area contains only the text area and not the non-text area.

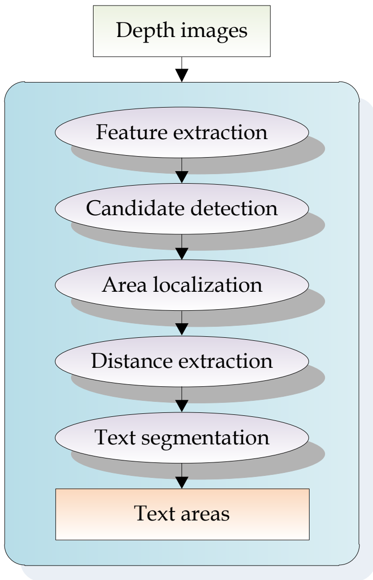

Figure 1. Overall flow chart of the proposed method

In Chapter 1, it is described the motivation and background of this study and a general overview. In Chapter 2, it is introduced the technique for extracting candidate character areas. In Chapter 3, it is described how to localize character areas. In Chapter 4, it is described the technique of segmenting localized text areas into text and background areas. In Chapter 5, it is explained how to verify candidate text regions. In Section 6, the experimental results are performed to examine the performance of the introduced text segmentation algorithm. In Section 7, the conclusion and future study directions are presented.

2. Acquisition of Candidate Text Areas

The candidate character region is segmented in the image using texture features. In order to extract the texture, we utilize the characteristic that high frequency components are large in the area where

March - April 2020
ISSN: 0193-4120 Page No. 4361 - 4368

characters exist in the image. First, after dividing an image into grids of 8×8 pixels, a discrete cosine transform (DCT) coefficient *co* like Equation (1) is extracted in grid units [11-13].

$$co_{uv} = \frac{1}{N} K_u K_v \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I_{xy} \times \cos\frac{\pi u(2x+1)}{2N} \cos\frac{\pi v(2y+1)}{2N} \quad (1)$$

In Equation (1), u and v denote frequencies in the transverse and longitudinal directions, N represents the number of pixels in the grid, and I means the gray-scale value of the input image. x and y represent the coordinates of the pixels in the grid. K is a coefficient. When u or v is 0, K is $1/\sqrt{2}$; otherwise K is 1.

In general, since the character area has a larger coefficient value of the transverse high-frequency component than the non-character region, the transverse high-frequency component of each grid is computed to remove blocks having low horizontal high-frequency components from each candidate block.

In this paper, only grids larger than the average of the transverse high-frequency components of all grids are selected as candidate blocks. Subsequently, the noise blocks are removed, and opening morphological operations [14-17] are performed using a 1×3 mask to connect the unconnected blocks. Then, blocks having a small longitudinal high-frequency component are deleted from the candidate character area. When the block is expanded using a 3×3 mask, the character area is enlarged. Therefore, we can restore the text area that was not detected because the high-frequency component is not sufficiently generated, including only the text area, the end of the text stroke, or only a very small region of the character stroke. Finally, a minimum enclosing rectangle (MER) corresponding to the detected character area is obtained. Then, the input image corresponding to the MER is acquired [18-20].

Since the candidate character area extracted in MER unit in the previous step is difficult to grasp the exact position of the string and also includes many non-character areas, the character area localization to minimize and extract to the area forming only the string should be performed. To this end, we utilize the property that a large amount of edges form a cluster near the text's stroke. In this study, we extract the edge map corresponding to the candidate character region using the Sobel edge mask [21-23].

Subsequently, the noise edges in the form of lines connected in the transverse and longitudinal directions are deleted. The edge map is then projected in the horizontal and vertical directions to produce a string of lines. To do this, we create an edge histogram and apply thresholds to determine if the edges are dense in a row or column. In addition, when detecting line-by-line strings, spaces should be considered. In general, the spacing in a text string is less than the height of the string.

Finally, character recognition is meaningless when the height of MER is less than or equal to the height where character recognition is not possible, and when the width of candidate text MER is smaller than the height.

3. Extraction of Text Areas

After localizing the text area, the localized text area is divided into background and character regions. Normally, there is a prior fact that the text is black and the non-text is white in the document. However, the subject image of this study does not have such prior knowledge. Therefore, in this paper, it is needed to distinguish between character and non-character in the character region, and to decide the gray-scale value for each of them. This procedure proceeds as follows.

First, an input image is converted into an intensity image, the Otsu's method [24-26] is used to obtain a best parameter value using histogram

March - April 2020
ISSN: 0193-4120 Page No. 4361 - 4368

segmentation, and then the segmented gray-scale image is converted using Equation (2). In Equation (2), GR(x, y) is an intensity image, and BI(x, y) is a binarized image. In addition, T_1 represents a threshold value.

$$BI(x, y) = \begin{cases} 255, & \text{if } GR(x, y) \geq T_1 \\ 0, & \text{if } GR(x, y) \prec T_1 \end{cases} \tag{2}$$

Second, two binarization maps are generated from the binarized image. Third, while scanning the image, the maximum inscribed rectangle (MIR) where the text area exists is extracted. Fourth, the average and variance of the MIR area are acquired using the length of one side of the MIRs acquired from the process so far. Sixth, the gray-scale value of each pixel is computed by comparing the maximum inscribed rectangle variance extracted from each map. If the variance of the map corresponding to black is small, the segmented image acquired in the first step is kept as it is, otherwise the binarized image is inversely transformed. This process generates an image B'(x, y) in which the gray-scale values of characters and non-characters pixels are determined. In this study, we set the gray-scale of character to 0 and the intensity of non-character to 255.

If the border contrast value of the letter is analogous to the contrast value of the non-character, the border of the letter is integrated into the non-character, so that the segmentation of the character and the non-character is carried out well. However, if the border contrast value of the letter is not similar to the text or the non-text, the segmentation of the character and the non-character does not proceed normally. Many existing methods have been proposed to solve this problem, but the results are not good. However, in this paper, we solve this problem by using distance information which is a three dimensional feature.

Candidate character regions extracted through the aforementioned character region localization process may still include non-character regions. Therefore, in this study, candidate text areas are verified by using 3D depth information and an artificial neural network [27-29], and only real text areas are selected except non-text regions. Although the conventional candidate character region verification algorithm using only artificial neural networks shows some accuracy, many misclassifications occur for candidate regions such as flowers, roofs, and wires, in which features analogous to the character regions are extracted.

In this paper, we verify the text area by combining three-dimensional depth information and an artificial neural network. First, a depth image is extracted through stereo matching from left and right input images. Second, after obtaining depth information of the pixel corresponding to the edge portion of the localized text area, the average and standard deviation of the extracted depth values are obtained. Third, since the depth values corresponding to the character areas are similar, the character areas with standard deviations or more are determined to be non-letter areas and are first removed. Fourthly, the text area is secondarily verified using a neural network-based learning. That is, the direction of the edge for the first-validated character area is divided into 16 directions, the size of the edge in each direction is taken as the input of the artificial neural network, and then non-character areas are secondarily removed by comparing with the value learned for the background area.

In this paper, to extract three-dimensional depth information, we apply the stereo matching algorithm [30-32] based on graph cut [33-35], which is known to show good performance recently among several existing methods. For this purpose, the image is converted into an intensity image, and the disparity information is extracted by matching the left and right images. The neural

March - April 2020
ISSN: 0193-4120 Page No. 4361 - 4368

network-based learning utilized in this study consists of 32 input nodes without bias nodes, one hidden layer with 28 nodes without bias nodes, and an output layer with two nodes. The feature used in the neural network-based learning is the size of the edge for the foreground and background of the candidate character area. Equations (3) and (4) are used for determining the direction and magnitude of the edge used for the feature.

$$ED_{direction}(x,y) = \tan^{-1}\frac{I(x,y+1)-I(x,y-1)}{I(x+1,y)-I(x-1,y)} \qquad (3)$$

$$ED_{magn}(x,y) = \sqrt{(I(x,y+1)-I(x,y-1))^2+(I(x+1,y)-I(x-1,y))^2} \qquad (4)$$

The directions of the edges calculated for each of the foreground and background of the candidate character region are divided into 16 directions to generate 32 histograms used as inputs of the neural network-based learning. The histograms 1 to 16 accumulate the magnitude of the edge obtained from the character region in bins along each direction. The histograms 17 to 32 accumulate the magnitude of the edge obtained from the non-character region in bins along each direction. Then, the corresponding histogram is divided by the total size of the character edge or the total size of the non-character edge.

4. Experimental Results

The laptop computer used for the testing consists of an Intel Pentium Core 2 Duo 2.66GHz CPU and 8GB of memory, and Microsoft Windows 10 was used as the operating system. In addition, Microsoft's Visual C ++ 2015 integrated development environment (IDE) was utilized to develop the character region extraction approach using the three-dimensional depth information suggested in this study. Additionally, a variety of stereoscopic color images were acquired and used to construct the character image dataset to be made use of in the experiment.

In this paper, in order to quantitatively compare the performance of the introduced character region extraction method, we used a correctness metric expressed as a percentage that shows the ratio of exactly localized strings that do not contain non-character areas to the total localized strings, as in Equation (5). In Equation (5), NO_{acc_text} represents the number accurately localized texts and NO_{local_text} denotes the number of localized texts.

$$M_{acc} = \frac{NO_{acc\_text}}{NO_{local\_text}} \times 100\ (\%) \qquad (5)$$

Fig. 2 quantitatively shows the accuracy of the conventional neural network method and the text area verification method using the proposed three-dimensional depth information. In general, when there is an artificially inserted background, the background area is simple, so that the binarization of the text area is performed accurately. However, in the case of text without artificial background, it is difficult to binarize the text and background in the existing method, but the proposed method can obtain relatively good results by using depth information.

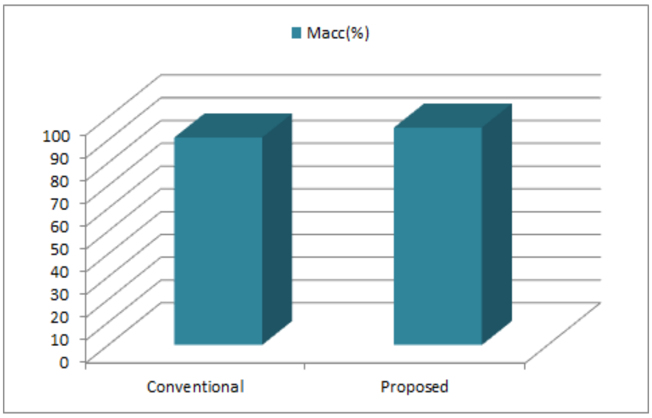

Figure 2. Performance evaluation

As can be seen in Fig. 2, the introduced algorithm improves the accuracy of character region verification by greatly reducing the extraction of non-character regions with the help of three-dimensional depth information.

March - April 2020
ISSN: 0193-4120 Page No. 4361 - 4368

In this study, text recognition was carried out by receiving the text area extracted in the previous step into the character recognizer. In this study, text recognition was carried out using READ IRIS Pro 10, which is known to have a good character recognition rate. As expected, the character recognition based on the proposed character segmentation method showed higher recognition rate than the conventional method.

5. Conclusion

In general, text information provides very important information for effectively understanding the content or meaning of an image. Therefore, there is a need for a study on accurately segmenting characters included in various images as a prerequisite of character recognition for indexing and searching multimedia materials. Therefore, in this study, we introduce a novel approach that effectively partitions the text region existing in three-dimensional stereoscopic images by combining texture and depth features.

First, candidate character regions expected to be located from the input image is extracted using texture information, and localization is performed to extract only an area in which a character string exists except the background region from candidate character regions. Then, the candidate text areas were verified using the depth information, and only actual text areas were segmented except non-character regions. Experimental results show that the introduced approach extracts the text area more robustly than the conventional two-dimensional text area segmentation method.

In the future, we will attempt to improve the performance of the text area extraction method using texture and depth data introduced in this study by testing more various types of stereoscopic color images. We will attempt to extract character regions from dynamic images that contain distorted character regions by applying artificial intelligence algorithm. In addition, it is planned to consolidate the algorithm by repeatedly tuning many arguments used in the proposed algorithm.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2019R1F1A1056475)

References

[1] Hashem ARA, Idris MYI, Ahmad AEBA. Comparative study of different binarization methods through their effects in characters localization in scene images. Data and Knowledge Engineering 2018;117:216-224.
[2] Zhang Y, Lai J, Yuen PC. Text string detection for loosely constructed characters with arbitrary orientations. Neurocomputing 2015;16830:970-978.
[3] Khare V, Shivakumara P, Chan CS, Lu T, Meng LK, Woon HH, et al. A novel character segmentation-reconstruction approach for license plate recognition. Expert Systems with Applications 2019;1311:219-239.
[4] Kanagarathinam K, Sekar K. Text detection and recognition in raw image dataset of seven segment digital energy meter display. Energy Reports 2019;5:842-852.
[5] Yi C, Tian Y. Text extraction from scene images by character appearance and structure modeling. Computer Vision and Image Understanding 2013;117:182-194.
[6] Qian X, Liu G, Wang H, Su R. Text detection, localization, and tracking in compressed video. Signal Processing: Image Communication 2007;22(9):752-768.
[7] Lyu MR, Song J, Cai M. A comprehensive method for multilingual video text detection, localization, and extraction. IEEE Transactions on Circuits and Systems for Video Technology 2005;15(2):243-255.
[8] Kim W, Kim C. A new approach for overlay text detection and extraction from complex video

March - April 2020
ISSN: 0193-4120 Page No. 4361 - 4368

scene. IEEE Transactions on Image Processing 2009;18(2):401-411.

[9] Golait D, Hubballi N. Detecting anomalous behavior in VoIP systems: a discrete event system modeling. IEEE Transactions on Information Forensics and Security 2017;12(3):730-745.

[10] Roy PP, Pal U, Llados J, Delalandre M. Multi-oriented touching text character segmentation in graphical documents using dynamic programming. Pattern Recognition. 2012;45(5):1972-1983.

[11] Ko HJ, Huang CT, Horng G, WANG SJ. Robust and blind image watermarking in DCT domain using inter-block coefficient correlation. Information Sciences 2020; 517:128-147.

[12] Ong SY, Li S, Wong KS, Tan KY. Fast recovery of unknown coefficients in DCT-transformed images. Signal Processing: Image Communication 2017;58:1-13.

[13] Parah SA, Sheikh JA, Loan NA, Bhat GM. Robust and blind watermarking technique in DCT domain using inter-block coefficient differencing. Digital Signal Processing 2016;53:11-24.

[14] Goyal B, Dogra A, Agrawal S, Sohi BS. Two-dimensional gray scale image denoising via morphological operations in NSST domain and bitonic filtering. Future Generation Computer Systems 2018;82:158-175.

[15] Parida P, Bhoi N. 2-D Gabor filter based transition region extraction and morphological operation for image segmentation. Computers and Electrical Engineering 2017;62:119-134.

[16] Hasan SMA, Ko K. Depth edge detection by image-based smoothing and morphological operations. Journal of Computational Design and Engineering. 2016;3:191-197.

[17] Borra SR, Reddy GJ, Reddy ES. Classification of fingerprint images with the aid of morphological operation and AGNN classifier. Applied Computing and Informatics 2018;14:166-176.

[18] Jiang N, Yang W, Duan L, Xu X, Liu Q. Acceleration of CT reconstruction for wheat tiller inspection based on adaptive minimum enclosing rectangle. Computers and Electronics in Agriculture 2012;85:123-133.

[19] Kwak E, Habib A. Automatic representation and reconstruction of DBM from LiDAR data using recursive minimum bounding rectangle. ISPRS Journal of Photogrammetry and Remote Sensing 2014;93:171-191.

[20] Zhang H, Tian X, Wang C, Wu F, Zhang B. Merchant vessel classification based on scattering component analysis for COSMO-SkyMed SAR images. IEEE Geoscience and Remote Sensing Letters 2013;10:1275-1279.

[21] Wang Z, Wang K, Yang F, Pan S, Han Y. Image segmentation of overlapping leaves based on Chan–Vese model and Sobel operator. Information Processing in Agriculture 2018;5:1-10.

[22] Furnari A, Farinella GM, Bruna AR, Battiato S. Distortion adaptive Sobel filters for the gradient estimation of wide angle images. Journal of Visual Communication and Image Representation 2017;46:165-175.

[23] Boustani B, Javaherian A, Nabi-Bidhendi M, Torabi S, Amindavar HR. Mapping channel edges in seismic data using curvelet transform and morphological filter. Journal of Applied Geophysics 2019;160:57-68.

[24] Merzban MH, Elbayoumi M. Efficient solution of Otsu multilevel image thresholding: A comparative study. Expert Systems with Applications 2019;116:299-309.

[25] Wang F, Pan X. Image segmentation for somatic cell of milk based on niching particle swarm optimization Otsu. Engineering in Agriculture, Environment and Food 2019;12(2):141-149.

[26] Sha C, Hou J, Cui H. A robust 2D Otsu's thresholding method in image segmentation. Journal of Visual Communication and Image Representation 2016;41:339-351.

[27] Castro FZD, Valle ME. A broad class of discrete-time hypercomplex-valued Hopfield neural networks. Neural Networks 2020;122:54-67.

[28] Santra B, Paul A, Mukherjee DP. Deterministic dropout for deep neural networks using composite random forest. Pattern Recognition Letters 2020;131:205-212.

[29] ZahediNasab R, Mohseni H. Neuroevolutionary based convolutional neural network with adaptive activation functions. Neurocomputing 2020;38114:306-313.

March - April 2020
ISSN: 0193-4120 Page No. 4361 - 4368

[30] Fu K, Xie Y, Jing H, Zhu J. Fast spatial-temporal stereo matching for 3D face reconstruction under speckle pattern projection. Image and Vision Computing 2019;85:36-45.
[31] Menant J, Gautier G, Pressigout M, Morin L, Nezan JF. An automatized method to parameterize embedded stereo matching algorithms. Journal of Systems Architecture 2017;80:92-103.
[32] Hamzah RA, Kadmin AF, Hamid MS, Ghani SFA, Ibrahim H. Improvement of stereo matching algorithm for 3D surface reconstruction. Signal Processing: Image Communication 2018;65:165-172.
[33] Zheng Q, Li H, Fan B, Wu S, Xu J. Integrating support vector machine and graph cuts for medical image segmentation. Journal of Visual Communication and Image Representation 2018;55:157-165.
[34] Peng Z, Qu S, Li Q. Interactive image segmentation using geodesic appearance overlap graph cut. Signal Processing: Image Communication 2019;78:159-170.
[35] Sun W, Dong E. Kullback-Leibler distance and graph cuts based active contour model for local segmentation. Biomedical Signal Processing and Control 2019;52:120-127.