

Robust Approach for Product Label Reading

Shradha Vitthalrao Chalikwar¹, Sinu Nambiar²

¹Savitribai Phule Pune University, Indira College of Engineering and Management, Pune, Maharashtra, India

²Professor, Savitribai Phule Pune University, Indira College of Engineering and Management, Pune, Maharashtra, India

Abstract: Existing system implements camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. Existing work propose an efficient and effective motion based method to define a region of interest (ROI) in the video by asking the user to shake the object. This method extracts moving object region by a mixture-of-Gaussians-based background subtraction method. In the extracted ROI, text localization and recognition are conducted to acquire text information. The existing method propose a novel text localization algorithm by learning gradient features of stroke orientations and distributions of edge pixels in an Adaboost model. In our proposed work our aim is to achieve the better accuracy using proposed algorithm. Present a novel image operator that seeks to find the value of stroke width for each image pixel, and demonstrate its use on the task of text detection in natural images. The proposed operator is local and data dependent, which makes it fast and robust enough to eliminate the need for multi-scale computation or scanning windows.

Keywords: Adaboost model, Text localization, OCR, Edge Detection, ROI

1. Introduction

In worldwide there are 314 million people visually impaired and 45 million are blind. In developed country for e.g. The U.S., in the 2008 national health interview survey reported that the 25.2 million adult Americans over 8% are blind or visually impaired. This number is increasing rapidly as the baby boomer generation ages. In Recent developments of computer vision, digital cameras, and portable computers make it easy to assist these individuals by developing camera-based products that combine computer vision technology with other existing commercial products such as optical character recognition (OCR) systems.

There are few systems that have some declaration for portable use, but they cannot handle the product labeling of system. For example the portable bar code reader designed to help blind people to identify product from product database it can enable users who are blind to access information about these product through speech. But in this process there is some limitation because it is very hard for blind user to find the position of bar code and to correctly point the bar code reader at the bar code. Therefore those situations handle by OCR software.

The OCR software provides the function of scanning and recognition of text and some have integrated voice output. However, these systems are generally designed for and perform best with document images with simple background, standard fonts, a small range of font sizes, and well organized characters. Most state-of-the-art OCR software cannot directly handle scene images with complex backgrounds. to extract text information from complex backgrounds with multiple and variable text patterns, propose a text localization algorithm that combines rule based layout analysis and learning-based text classifier training, which define novel feature maps based on stroke orientations and edge distributions. And generate representative and discriminative text features to distinguish text characters from background outliers.

2. Related Work

2.1 Literature Review

1. Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons

In this paper, this method extracts moving object region by a mixture-of-Gaussians-based background subtraction method. In the extracted ROI, text localization and recognition are conducted to acquire text information [1]. To automatically localize the text regions from the object ROI, we propose a novel text localization algorithm by learning gradient features of stroke orientations and distributions of edge pixels in an Ad boost model. Text characters in the localized text regions are then binarized and recognized by off-the-shelf optical character recognition software. The recognized text codes are output to blind users in speech [1].

Rule-based algorithms apply pixel-level image processing to extract text information from predefined text layouts such as character size, aspect ratio, edge density, character structure, color uniformity of text string, etc. It Analyzed edge pixel density and performed global binarization to obtain text regions. In color-based text segmentation is performed through a Gaussian mixture model for calculating a confidence value for text regions. This type of algorithm tries to define a universal feature descriptor of text [2].

Learning-based algorithms build the model text structure and extract representative text features and text classifiers. Chen and Yuille [4] presented five types of Haarbased block patterns to train text classifiers in Adaboost learning. We have previously developed rule-based algorithms to extract text from scene images.

2. Multiscale Edge-Based Text Extraction from Complex Images

Text that appears in images contains important and useful information. Detection and extraction of text in images have been used in many applications. Here, propose multiscale edge-based text extraction algorithm, which can

automatically[5] detect and extract text in complex images. It is robust with respect to the font size, style, color, orientation, and alignment of text and can be used in a large variety of application fields, such as mobile robot navigation, vehicle license detection and recognition, object identification, document retrieving, page segmentation, etc[5].

Advantage

- It presents an effective and robust general-purpose text detection and extraction algorithm, which can automatically detect and extract text from complex background images.
- It can handle both printed document and scene text images.
- Not sensitive to image color/intensity, robust with respect to font, sizes, orientations, alignment, uneven illumination, perspective and reflection effects.

Disadvantage

- Method only analyzes text blocks.

3. Local Features based Text Detection Techniques in Document Images

Video text information plays an important role in semantic-based video analysis, indexing and retrieval. It is observed that the detection of texts in video remains as a challenging task due to its complex varying conditions. In this paper present a study on local features based text detection in document images and more focus is provided for text detection based on Laplacian method. In addition, a comparative study of the Laplacian method[6] with novel text detection and localization method based on Corner response and Multi scale edge based method for video text detection is made[6].

Advantages

- Text data present in images and video contain useful information for automatic annotation, indexing and structuring of images.
- It helps to identify the candidate text regions and the edge information serves to determine the accurate boundary of each text block.
- It reduces noise in the feature extraction stage.

Disadvantages

- Variations of text due to differences in size, style, orientation and alignment as well as low image contrast and complex background make the problem of automatic text detection extremely challenging.
- Unable to detect texts of different fonts, contrast and backgrounds.

4. Camera based Assistive Text Reading System using Gradient and Stroke Orientation for Blind Person

Propose an accurate and robust method for detecting texts in object containing images. Stationary wavelet transform is applied to identify the text regions from the image. Text regions and non-text regions are separated by the connected component analysis. The non-text regions and other region in the image are removed. Character candidates are grouped into text candidates using Edge Distribution & stroke and

gradient orientation maps. Canny edge detection algorithm is proposed to extract the feature from the image. The recognized text in the images are classified and converted to speech using text to speech converter using MATLAB toolbox and produced as audio to the blind people.

Advantage

- It improves quality of an image.

Disadvantage

- It cannot change the font style and cannot detect text in different structure.

3. Proposed Work

In proposed system the paper represents a prototype system of assistive text reading. From Fig., the system framework consists of three functional components:

- 1) Scene Capture,
- 2) Data Processing, and
- 3) Audio Output.

a. Object Region Detection

The hand-held object appears in the camera view, employ a camera with a reasonably wide angle in our prototype system. However, this may result in some other extraneous but perhaps text-like objects appearing in the camera view. To extract the hand-held object of interest from other objects in the camera view, we ask users to shake the hand-held objects containing the text they wish to identify and then employ a motion-based method to localize the objects from cluttered background. To detect moving objects in a dynamic scene, many adaptive BGS techniques have been developed. This Gaussian-mixture-model based method is robust to slow lighting changes, but cannot handle complex foregrounds and quick lighting changes. Further improved the multiple Gaussian-mixture based BGS method to better define foreground while remove background objects. First, texture information is employed to remove false positive foreground areas. A texture similarity measure is defined to evaluate whether the detected foreground motion is caused by lighting change or moving object.

b. Text Localization:

• The Stroke Width Transform:

The Stroke Width Transform (SWT) is a local image operator which computes per pixel the width of the most likely stroke containing the pixel. The output of the SWT is an image of size equal to the size of the input image where each element contains the width of the stroke associated with the pixel. In Figure 1(a). We define a stroke to be a contiguous part of an image that forms a band of a nearly constant width.

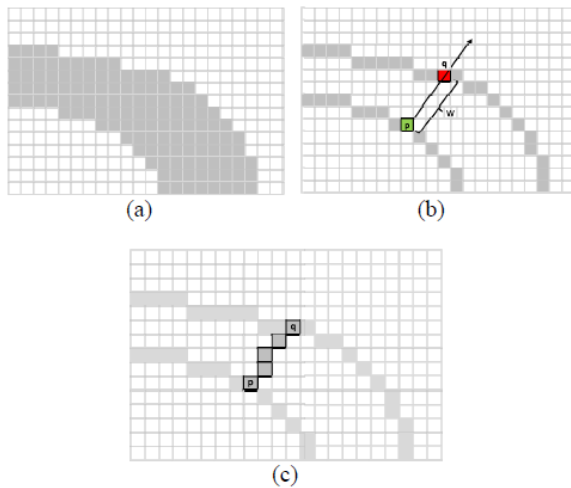


Figure 1: SWT.

The initial value of each element of the SWT is set to ∞ . In order to recover strokes, first compute edges in the image using canny edge detector. After that, a gradient direction dp of each edge pixel p is considered (Fig.1 b). If p lies on a stroke boundary, then dp must be roughly perpendicular to the orientation of the stroke. Follow the ray $r=p+n \cdot dp, n>0$ until another edge pixel q is found. consider then the gradient direction dq at pixel q . If dq is roughly opposite to dp ($d = -dp \pm \pi/6$), each element S of the SWT output image corresponding to the pixels along the segment $[p,q]$ is assigned the width $\|p-q\|$ unless it already has a lower value from Fig. 2a.

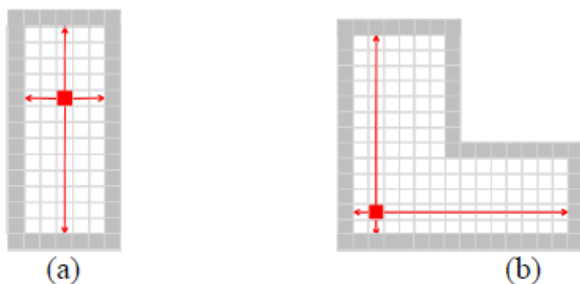


Figure 2: Otherwise, if the matching pixel q is not found, or if dq is not opposite to dp , the ray is discarded.

As shown in Fig. 2b, the SWT values in more complex situations, like corners, will not be true stroke widths after the first pass described above. Therefore, we pass along each non-discarded ray again, compute median SWT value m of all its pixels, and then set all the pixels of the ray with SWT values above m to be equal.

The output of the SWT is an image where each pixel contains the width of stroke it belongs to the next step of the algorithm is to group these pixels into letter candidates. We now need to identify components that may contain text. For this there are some rules.

The first we perform is to compute the variance of the stroke width within each connected component and reject the ones whose variance is too big. At the last the components whose size is too small or too large may be ignored.

c. Grouping Letters into text line

To further increase the reliability of the algorithm, we continue a step forward to consider groups of letters finding such groups is a significant filtering mechanism as single letters do not usually appear in images and this reasoning allows us to remove randomly scattered noise. An important cue for text is that it appears in a linear form. Including this reasoning proves to be both straightforward and valuable. For example, a lamp post next to a car wheel would not be mistaken for the combination of letters "O" and "I" as the Post is much higher than the wheel. We consider each pair of letter candidates for the possibility of belonging to the same text line. Two letter candidates should have similar stroke width (ratio is Not exceeding 2.0 (due to the difference between capital and lower case letters). The distance between letters must not exceed three times the width of the wider one. Additionally, average colors of candidates for pairing are compared, as letters in the same word are typically expected to be written in the same color.

At the next step of the algorithm, the candidate pairs determined above are clustered together into chains. Initially, each chain consists of a single pair of letter candidates. The process ends when no chains can be merged. Each produced chain of sufficient length (at least 3 letters in our experiments) is considered to be a text line. Finally, text lines are broken into separate words, using a heuristic that computes a histogram of horizontal distances between consecutive letters and estimates the distance threshold that separates intra-word letter distances from inter-word letter distances.

d. Text Recognition And Audio Output

Text recognition is performed by off-the-shelf OCR. A text region labels the minimum rectangular area for the accommodation of characters inside it, so the border of the text region contacts the edge. Thus, each localized text region is enlarged by enhancing the height and width by 10 pixels, respectively, and then, we use Otsu's method to perform binarization of text regions, where margin areas are always considered as background. We test both open- and closed-source solutions that allow the final stage of conversion to letter codes (e.g. Omni Page). The recognized text codes are recorded in script files. Then, we employ the Microsoft Speech Software Development Kit to load these files and display the audio output of text information. Blind users can adjust speech rate, volume, and tone according to their preferences.

4. Conclusion

We propose an algorithm for the text detection from the images in our work. The proposed work contain important phases like object detection, text localization and audio output. The proposed system will be more efficient and effective for the use by the blind persons. In our approach we use stroke width transformation method for text localization. We use Otsu's method to perform binarization of text regions. We use script for the conversion of text into audio output.

5. Acknowledgement

Authors are thankful to the guide. Prof. Sinu Nambiar for continuous guidance and support. I also express our thanks for those who contributed directly or indirectly to complete this work.

References

- [1] Chucui Yi, Student Member, IEEE, YingliTian, Senior Member, IEEE, and Aries Arditi, "Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind Persons", IEEE/ASME TRANSACTIONS ON MECHATRONICS, VOL. 19, NO. 3, JUNE 2014.
- [2] World Health Organization. (2009). 10 facts about blindness and visual impairment [Online]. Available: www.who.int/features/factfiles/blindness/blindness_facts/en/index.html
- [3] Advance Data Reports from the National Health Interview Survey (2008). [Online]. Available: http://www.cdc.gov/nchs/nhis/nhis_ad.htm
- [4] International Workshop on Camera-Based Document Analysis and Recognition (CBDAR 2005, 2007, 2009, 2011). [Online]: Available: <http://www.m.cs.osakafu-u.ac.jp/cbdar2011/>
- [5] Xiaoqing Liu and JagathSamarabandu, "Multiscale Edge-Based Text Extraction From Complex Images."
- [6] M SharmilaKumari and Akshatha, "Local Features based Text Detection Techniques in Document Images."
- [7] X. Chen and A. L. Yuille, "Detecting and reading text in natural scenes," in Proc. Comput. Vision Pattern Recognit., 2004, vol. 2, pp. II-366–II-373.
- [8] X. Chen, J. Yang, J. Zhang, and A. Waibel, "Automatic detection and recognition of signs from natural scenes," IEEE Trans. Image Process., vol. 13, no. 1, pp. 87–99, Jan. 2004.