ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

ASSISTIVE PRODUCT LABEL READING FROM HAND HELD OBJECTS AND CLOTHING PATTERN RECOGNITION FOR VISUALLY IMPAIRED PEOPLE

J.STELLA MARY[1] andJ.JASMINE CHRISTINA[2]
[1]*PG student, Department of ECE, Vivekanandhacollege of engineering for women, Tiruchengode.*
[2]*PGstudent,Department of ECE, Vivekanandhacollege of engineering for women, Tiruchengod.*

Abstract- We propose a camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. To isolate the object from cluttered backgrounds or other surrounding objects in the camera view, we first propose an efficient and effective motionbased 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. 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 Adaboostmodel. Text characters in the localized text regions are then binarized and recognized by off-the-shelf optical character recognition software. recognized text codes are output to blind users in speech. We explore user interface issues and assess robustness of the algorithm in extracting and reading text from different objects with complex backgrounds.

***Index Terms*—Assistive devices, blindness, distribution of edge pixels, hand-held objects, optical character recognition (OCR), stroke orientation, text reading, text region localization.**

I. INTRODUCTION

OF the 314 million visually impaired people worldwide, 45million are blind [1]. Even in a developed country like the U.S., the 2008 National Health Interview Survey reported that an estimated 25.2 million adult Americans (over 8%) are blind or visually impaired [2]. This number is increasing rapidly as the baby boomer generation ages. Recent developments in computer vision, digital cameras, and portable computers make it feasible to assist these individuals by developing camera-based products that combine computer vision technology with other existing commercial products such optical character recognition (OCR) systems. Reading is obviously essential in today's society. Printed text is everywhere in the form of reports, receipts, bank statements, restaurant menus, classroom handouts, product packages, instructions on medicine bottles, etc. Andwhile optical aids, video magnifiers, and screen readers can help blind users and those with low vision to access documents, there are few devices that can provide good access to common hand-held objects such as product packages, and objects printed with text such as prescription medication bottles. The ability of people who are blind or have significant visual impairments to read printed labels and product packages will enhance independent living and foster economic and social self-sufficiency.

Today, there are already a fewsystems that have some promise for portable use, but they cannot handle product labeling. For example, portable bar code readers designed to help blind people identify different products in an extensive product database can enable users who are blind to access information about these products [2] through speech and braille. But a big limitation is that it is very hard for blind users to find the position of the bar code and to correctly point the bar code reader at the bar code. Some reading-assistive systems such as pen scanners might be employed in these and similar situations. Such systems integrate OCR software to offer 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 backgrounds, standard fonts, a small range of font sizes, and well-organized characters rather than commercial product boxes with multiple decorative patterns. Most state-of-the-art OCR software cannot directly handle scene images with complex backgrounds.

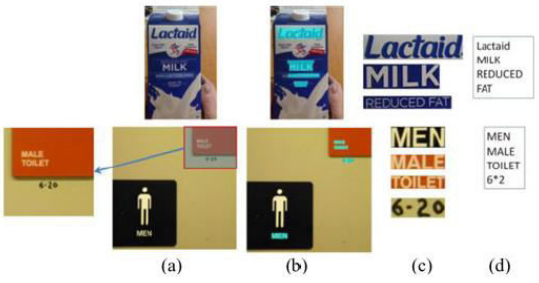
ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

A number of portable reading assistants have been designed specifically for the visually impaired [2]

Fig. 1. Examples of printed text from hand-held objects with multiple colors, complex backgrounds, or nonflat surfaces.

Mobile accurately reads black print on a white background, but has problems recognizing colored text or text on a colored background. It cannot read text with complex backgrounds, text printed on cylinders with warped or incomplete images (such as soup cans or medicine bottles).

Furthermore, these systems require a blind user to manually localize areas of interest and text regions on the objects in most cases. Although a number of reading assistants have been designed specifically for the visually impaired, to our knowledge, no existing reading assistant can read text from the kinds of challenging patterns and backgrounds found on many everyday commercial products. As shown in Fig. 1, such text information can appear in multiple scales, fonts, colors, and orientations.



(a)camera capture image,(b) Localized text regions(c) Text regionscropped from image. (d) Text codes recognized by OCR. Text at the top-right corner of bottom image is shown in a magnified callout.

Our proposed algorithm can effectively handle complex background and multiple patterns, and extract text information from both hand-held objects and nearby signage, as shown in Fig. 2. In assistive reading systems for blind persons, it is very challenging for users to position the object of interest within the center of the camera's view. As of now, there are still no acceptable solutions. We approach the problem in stages. To make sure the hand-held object appears in the camera view, we use a camera with sufficiently wide angle to accommodate users with only approximate aim. This may often result in other text objects appearing in the camera's view (for example, while shopping at a supermarket). To extract the hand-held object from the camera image, we develop a motion-based method to obtain a region of interest (ROI) of the object. Then, we perform text recognition only in this ROI. It is a challenging problem to automatically localize objects and text ROIs from captured images with complex backgrounds, because text in captured images is most likely surrounded by various background outlier "noise," and text characters usually appear in multiple scales, fonts, and colors. For the text orientations, this paper assumes that text strings in scene images keep approximately horizontal alignment. Many algorithms have been developed for localization of text regions in scene images. We divide them into two categories: rule-based and learning-based. 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. Phan *et al.* [4] analyzed edge pixel density with the Laplacian operator and employed maximum gradient differences to identify text regions.

This type of algorithm tries to define a universal feature descriptor of text. Learning-based algorithms, on the other hand, model text structure and extract representative text features to build text classifiers. Chen and Yuille [4] presented five types of Haarbased block patterns to train text classifiers in an Adaboost learning model. Kim *et al.* [1] considered text as a specific texture and analyzed the textural features of characters by a support vector machine (SVM) model. Kumar *et al.* [3] used globally matched wavelet filter responses of text structure as features. Ma *et al.* [5] performed classification of text edges by using histograms of oriented gradients and local binary patterns as local features on the SVM model. Shi *et al.* [5] employed gradient and curvature features to model the grayscale curve for handwritten numeral recognition under a Bayesian discriminant function. In our

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

research group, we have previously developed rule-based algorithms to extract text from scene images[3]–[5]. A survey paper about computer-vision-based assistive technologies to help people with visual impairments can be found in [5]. In solving the task at hand, to extract text information from complex backgrounds with multiple and variable text patterns, we here propose a text localization algorithm that combines rulebased layout analysis and learning-based text classifier training, which define novel feature maps based on stroke orientations and edge distributions. These, in turn, generate representative and discriminative text features to distinguish text characters from background outliers.

II.FRAMEWORK AND ALGORITHM OVERVIEW

This paper presents a prototype system of assistive text reading. As illustrated in Fig. 3, the system framework consists of three functional components: scene capture, data processing, and audio output. The scene capture component collects scenescontaining objects of interest in the form of images or video. In our prototype, it corresponds to a camera attached to a pair of sunglasses.

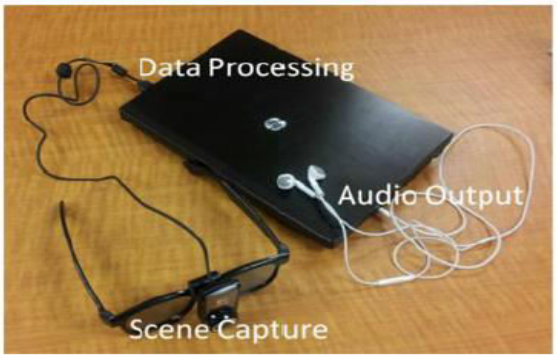
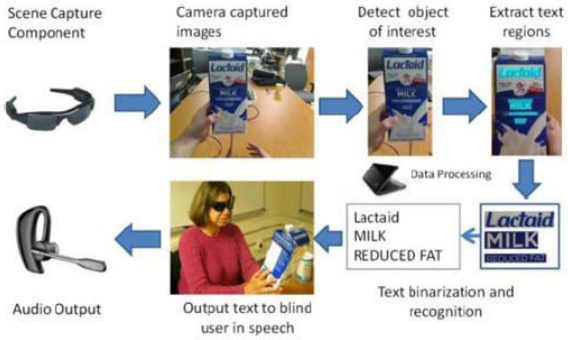


Fig. 3. Snapshot of our demo system, including three functional componentsfor scene capture, data processing, and audio output.

. The data processing component is used for deploying our proposed algorithms, including 1) object-of-interest detection to selectively extract the image of the object held by the blind user from the cluttered background or other neutral objects in the camera view; and 2) text localization to obtain image regions containing text, and text recognition to transform image-based text information into readable codes.We use aminlaptop as the processing device in our current prototype system. The audio output component is to inform the blind user of recognized text codes. A Bluetooth earpiece with mini microphone is employed for speech output.



This simple hardware configuration ensures the portability of the assistive text reading system. Fig. 4 depicts a work flowchart of the prototype system. A frame sequence V is captured by a camera worn by blind users, containing their hand-held objects and cluttered background. To extract text information from the objects, motionbased object detection is first applied to determine the user's object of interest S by shaking the object while recording video

$$S = \frac{1}{|V|}\sum_{i}\mathcal{R}(V_i, B)$$

where Vi denotes the ith frame in the captured sequence, $|V|$ denotes the number of frames, B denotes the estimated background from motion-based object detection, and R represents the calculated foreground object at each frame. The object of interest is localized by the average of foreground masks (see details in Section III). Next, our novel proposed text localization algorithm is applied to the object of interest to extract text regions. At first, candidate text regions are generated by layout analysis of color uniformity and horizontal alignment

$$X^C = \text{argmax}_{s\in S} L(s)$$

where $L(\cdot)$ denotes the suitability responses of text layout and XC denotes the candidate text regions

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

from object of interest *S*. Then, a text classifier is generated from a Cascade-Adaboost learning model, by using stroke orientations and edge distributions of text characters as features (see details in Section IV).

$$X = H\left[X^C\right] = H\left[\text{argmax}_{s \in S} L(s)\right]$$

where *H* denotes the Cascade-Adaboost classifier and *X* denotes the localized text regions.

After text region localization, off-the-shelf OCR is employed to perform text recognition in the localized text regions. The recognized words are transformed into speech for blind users. Our main contributions embodied in this prototype system are: 1) a novelmotion-based algorithm to solve the aiming problem for blind users by their simply shaking the object of interest for a brief period; 2) a novel algorithm of automatic text localization to extract text regions from complex background and multiple text patterns; and 3) a portable camera-based assistive framework to aid blind persons reading text from hand-held
objects. Algorithms of the proposed system are evaluated over images captured by blind users using the described techniques.

III. AUTOMATIC TEXT EXTRACTION

As shown in Fig. 5, we design a learning-based algorithm for automatic localization of text regions in image. In order to handle complex backgrounds, we propose two novel feature maps to extracts text features based on stroke orientations and edge distributions, respectively. Here, stroke is defined as a uniform region with bounded width and significant extent. These feature maps are combined to build an Adaboost based text classifier.

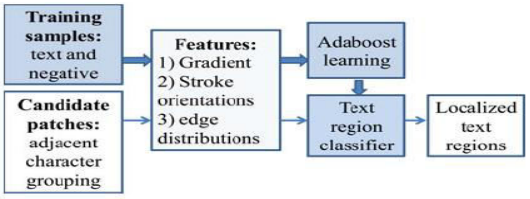


Fig. 5. Diagram of the proposed Adaboost learning-based text region localization algorithm by using stroke orientations and edge distributions.

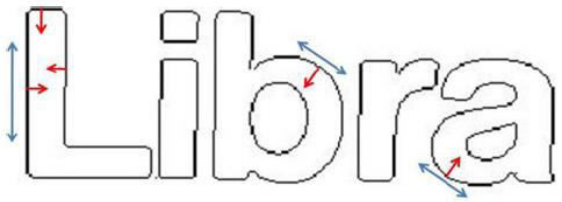


Fig. 6. Sample of text strokes showing relationships between stroke orientations
and gradient orientations at pixels of stroke boundaries.

Blue arrows denote the stroke orientations at the sections and red arrows denote the gradient orientations at pixels of stroke boundaries

A. TEXT STROKE ORIENTATION

Text characters consist of strokes with constant or variable orientation as the basic structure. Here, we propose a new type of feature, stroke orientation, to describe the local structure of text characters. From the pixel-level analysis, stroke orientation is perpendicular to the gradient orientations at pixels of stroke boundaries, as shown in Fig. 6. To model the text structure by stroke orientations, we propose a new operator to map a gradient feature of strokes to each pixel.

It extends the local structure of a stroke boundary into its neighborhood by gradient of orientations. We use it to develop a feature map to analyze global structures of text characters. Given an image patch *I*, Sobel operators in horizontal and vertical derivatives are used to calculate two gradient maps *Gx* and*Gy* , respectively. The synthesized gradient map is calculated as *G* = _*G*2*x* + *G*2*y*_1⁄2 . The Canny edge detector is applied on *I* to calculate its binary edge map *E*.

We set this threshold because the text patches in our experiments are all normalized into height 48 pixels and width 96 pixels, and the stroke width of text characters in these normalized patches mostly does not exceed 36. If the distance is greater than 36, pixel *p*0 would be located at background region far away from text character.

In the range, we select the edge pixel *pe* with the minimum Euclidean distance from *p*0 . Then, the pixel *p*0 is labeled with gradient orientation at pixel *pe* from gradient maps by

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

$$p_e = \underset{p \in P}{\operatorname{argmin}}\, d(p, p_0)$$

$$S(p_0) = \Upsilon(\arctan(G_y(p_e), G_x(p_e)))$$

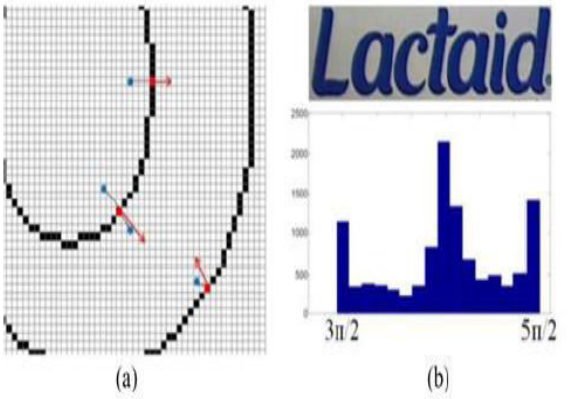


Fig.7 (a) Example of stroke orientation label. The pixels denoted by blue points are assigned the gradient orientations (red arrows) at their nearest edge pixels, denoted by the red points. (b) 210×54 text patch and its 16-bin histogram of quantized stroke orientations. where $P = \{p|p \in R(p0)$, p is edge pixel$\}$. The stroke orientation calculated from arctan will be in the range $(-\pi/2, \pi/2]$. To distinguish the pixels labeled with stroke orientation 0 and the unlabeled pixels also with value 0, Y shifts the stroke orientations one period forward into the range $(3\pi/2, 5\pi/2]$, which removes the value 0 from the range of stroke orientations. A stroke orientation map $S(p)$ is output by assigning each pixel the gradient orientation at its nearest edge pixel, as shown in Fig. 8(a).

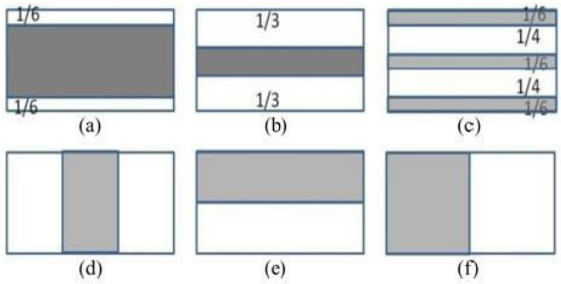


Fig. 10. Block patterns based on [4].

For the block patterns with more than two subregions [see Fig. 10(a)–(d)], the other metric of feature response is the absolute difference between the mean of pixel values in white regions and the mean of pixel values in black regions. Thus, we obtain 6 + (6 − 2) = 10 feature values through the six block patterns and two metrics from each feature map. The "integral image" algorithm is used in these calculations [3]. From the 18 feature maps (3 gradient maps, 14 stroke orientation maps, and 1 edge distribution map), a training sample can generate a feature vector of 180 dimensions as (3). We compute feature vectors for all the 51 234 samples in the training set. By using feature vector $\boldsymbol{fi}$ of the ith sample as the ith column, a feature matrix $\boldsymbol{F}$ is obtained by (2)

$$\boldsymbol{f}^i = \left[f_1^i, f_2^i, \ldots, f_{180}^i\right]^T$$

$$\boldsymbol{F} = \left[\boldsymbol{f}^1, \boldsymbol{f}^2, \ldots, \boldsymbol{f}^t, \ldots, \boldsymbol{f}^{51234}\right].$$

The 180 × 51 234 feature matrix is used for learning a text classifier in a Cascade-Adaboost model. A row of the feature matrix records feature responses of a certain block pattern and a certain feature map on all training samples. In the process of Adaboost learning, weak classifier is defined as r, Tr, ρ. The three parameters denote the rth row of feature matrix ($1 \le r \le 180$), a threshold of the rth row Tr , and polarity of the threshold $\rho \in \{-1, 1\}$. In each row r, linearly spaced threshold values are sampled in the domain of its feature values by

$$T_r \in \left\{T | T = f_r^{\min} + \frac{1}{N_T}\left(f_r^{\max} - f_r^{\min}\right) t\right\}$$

where NT represents the number of thresholds, $f$min r and $f$max r represent the minimum and maximum feature value of the rth row, respectively, and t is an integer ranging from 1 to NT . We set NT = 300 in the learning process. Thus, there are in total 180 × 2 × 300 = 108 000 weak classifiers denoted as$H$.When a weak classifier r, ρ, Tr is applied to a sample with corresponding feature vector $\boldsymbol{f}$ = [f1, . . . , fr , . . . , f180]T , if $\rho fr \ge \rho Tr$, it is classified as a positive sample; otherwise, it is classified as a negative sample. The Cascade-Adaboost classifier has proved to be an effective machine learning algorithm in real-time face detection [3].

The training process is divided into several stages. In each stage, a stage-specific Adaboost classifier is learned from a training set, which

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

consists of all positive samples and the negative samples incorrectly classified by previous Adaboost classifiers at this stage. We refer to this as a stage-Adaboost classifier in the following paragraphs. The learning process based on the Adaboost model [4] at each stage is as follows: 1) The set of m samples ($x1$, $y1$) , ($x2$, $y2$) , . . . , (xm, ym) is given, where $xi \in X$ denotes feature vector and $yi \in \{-1, 1\}$ denotes ground truth. Each sample i is assigned a weight Di , which is initialized to be $1/m$. 2) In the tth iteration, we select the optimized weak classifier ht from the set of weak classifiers H, such that ht = argmin$h \in H$ m i=1 $Diyih$ (xi), and calculate εt = m i=1 $D(i)$ · (yi _= h (xi)) and αt = 0.5 ln((1 − εt)/εt). 3) Update the sample weights by Di := Diexp(−yih(xi)). 4) Start the next iteration from step (2) until all the samples are correctly classified or the maximum number of iterations is reached. 5) The optimized weak classifiers are combined into a stage-Adaboost classifier as $H(x) = t$ αtht (x). In the end, all the stage-Adaboost classifiers are cascaded into the final Cascade-Adaboost classifier. When a test image patch is input into the final classifier, it is classified as a text patch if all the cascaded stage-Adaboost classifiers determine it is a positive sample, and otherwise, it is classified as a nontext patch. In the learning process, each stage-Adaboost classifier ensures that 99.5% of positive samples are correctly classified, while 50% of negative samples are correctly classified. Thus, a testing sample with positive ground truth will have a (0.995)T chance of correct classification, where T represents the total number of stage-Adaboost classifiers.

IV. TEXT RECOGNITION AND AUDIO OUTPUT

Text recognition is performed by off-the-shelf OCR prior to output of informative words from the localized text regions. 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 boundary of the text character. However, our experiments show that OCR generates better performance if text regions are first assigned proper margin areas and binarized to segment text characters from background. Thus, each localized text region is enlarged by enhancing the height and width by 10 pixels, respectively, and then, we use Otsu's method [1] to perform binarization of text regions, wheremargin 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. OmniPage, Tesseract, ABBYReader). The recognized text codes are recorded in script files. Then, we employ the Microsoft Speech Software Development Kit toload these files and display the audio output of text information. Blind users can adjust speech rate, volume, and tone according to their preferences.

V. EXPERIMENTS

A. Datasets

Two datasets are used to evaluate our algorithm. First, the ICDARRobustReading Dataset [1], [4] is used to evaluate the proposed text localization algorithm. The ICDAR-2003 dataset contains 509 natural scene images in total. Most images contain indoor or outdoor text signage. The image resolutions range from 640 × 480 to 1600 × 1200. Since layout analysis based on adjacent character grouping can only handle text strings with three or more character members, we omit the images containing only ground truth text regions of less than three text characters.

Thus, 488 images are selected from this dataset as testing images to evaluate our localization algorithm. To further understand the performance of the prototype system and develop a user-friendly interface, following Human Subjects Institutional Review Board approval, we recruited ten blind persons to collect a dataset of reading text on hand-held

Fig. 12. Examples of blind persons capturing images of the object in their hands objects.

The hardware of the prototype system includes a Logitech web camera with autofocus, which is secured to the nose bridge of a pair of sunglasses.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

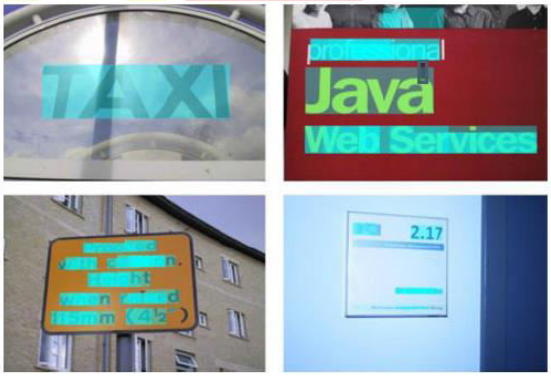
The camera is connected to an HP mini laptop by a USB connection. The laptop performs the processing and provides audio output. In order to avoid serious blocking or aural distraction, we would choose a wireless"open" style Bluetooth earpiece for presenting detection results as speech outputs to the blind travelers in a full prototype implementation.The blind user wore the camera/sunglasses to capture the image of the objects in his/her hand, as illustrated in Fig. 12.

B. Evaluations of Region Localization Text

Text classification based on the Cascade-Adaboost classifier plays an important role in text region localization. To evaluate the effectiveness of the text classifier, we first performed a group of experiments on the dataset of sample patches, in which the patches containing text are positive samples and those without text are negative samples. These patches are cropped from natural scene images in ICDAR-2003 and ICDAR-2011 Robust Reading Datasets. Each patch was assigned a prediction score by the text classifier; a higher score indicates a higher probability of text information. We define the true positive rate as the ratio of correct positive predictions to the total number of positive samples. Similarly, the false positive rate is the ratio of correct positive predictions to the total number of positive predictions.

This characteristic is compatible with the design of our blindassistive framework, in which it is useful to filter out extraneous background outliers and keep a conservative standard for what constitutes text. Next, the text region localization algorithm was performed on the scene images of ICDAR-2003 Robust Reading Datasetto identify image regions containing text information.

Some example results of text localization on the ICDAR-2003 robust reading dataset, and the localized text regions are marked in blue. It shows that our algorithm can localize multiple text labels in indoor and outdoor environments.



Figs. 14,15, and 16(a) depict some results showing the localization of text regions, marked by blue rectangular boxes.

For a pair of text regions, match score is estimated by the ratio between the intersection area and the mean area of the union of the two regions. Each localized (ground truth) text region generates a maximum match score from its best matched ground truth (localized) text region. *Precision* is the ratio of total match score to the total number of localized regions. It estimates the false positive localized regions. *Recall* is the ratio between the total match score and the total number of ground truth regions. It estimates themissing text regions.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Special Issue 13, April 2016

VI. CONCLUSION AND FUTURE WORK

In this paper, we have described a prototype system to read printed text on hand-held objects for assisting blind persons. In order to solve the common aiming problem for blind users, we have proposed a motion-based method to detect the object of interest, while the blind user simply shakes the object for a couple of seconds. This method can effectively distinguish the object of interest from background or other objects in the camera view. To extract text regions from complex backgrounds,we have proposed a novel text localization algorithm based on models of stroke orientation and edge distributions. The corresponding feature maps estimate the global structural feature of text at every pixel. Block patterns project the proposed feature maps of an image patch into a feature vector. Adjacent character grouping is performed to calculate candidates of text patches
prepared for text classification. An Adaboost learning model is employed to localize text in camera-based images. Off-the-shelf OCR is used to perform word recognition on the localized text regions and transform into audio output for blind users. Our future work will extend our localization algorithm to process text strings with characters fewer than three and to design
more robust block patterns for text feature extraction. We will also extend our algorithm to handle nonhorizontal text strings. Furthermore, we will address the significant human interface
issues associated with reading text by blind users.

REFERENCES

[1] X. Yang, Y. Tian, C. Yi, and A. Arditi, "Context-based indoor objectdetection as an aid to blind persons accessing unfamiliar environments,"in *Proc. ACM Multimedia*, 2010, pp. 1087–1090.

[2] X. Yang, S. Yuan, and Y. Tian, "Recognizing clothes patterns for blind people by confidence margin based feature combination," in *Proc. ACM Multimedia*, 2011, pp. 1097–1100.

[3] C. Yi and Y. Tian, "Text string detection from natural scenes by structure based partition and grouping," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2594–2605, Sep. 2011.

[4] C. Yi and Y. Tian, "Assistive text reading from complex background for blind persons," in *Proc. Int. Workshop Camera-Based Document Anal. Recognit.*, 2011, vol. LNCS-7139, pp. 15–28.

[5] C. Yi and Y. Tian, "Text detection in natural scene images by stroke gabor words," in *Proc. Int. Conf. Document Anal. Recognit.*, 2011, pp. 177–181.