Headline Based Text Extraction from Outdoor Images

Ranjit Ghoshal[1], Anandarup Roy[2],
Tapan Kumar Bhowmik[3], and Swapan K. Parui[2]

[1] St. Thomas' Collage of Engineering and Technology
Kolkata- 700023, India
ranjit.ghoshal@rediffmail.com
[2] CVPR Unit
Indian Statistical Institute, Kolkata, India
roy.anandarup@gmail.com,swapan@isical.ac.in
[3] Faculty of Mathematics and Natural Sciences,
University of Groningen, Netherlands
tkbhowmik@ai.rug.nl

Abstract. The goal of this article is to design an effective scheme for extraction of Bangla/Devnagari text from outdoor images. We first segment a color image using fuzzy c-means algorithm. In Bangla/Devnagari script, text may be attached/unattached to the headlines. Hence, after segmentation, headlines are detected from each connected components using morphology. Now, the components attached or close to the detected headlines are separated. Further by applying certain shape and position based purification we could distinguish text and non text. Our experiments on a dataset of 100 outdoor images containing Bangla and/or Devnagari text reveals satisfactory performance.

1 Introduction

With the increasing popularity of digital cameras attached with various handheld devices (mobile phones, PDAs etc), many new computational challenges have gained significance. Extraction and recognition of texts from outdoor images captured by such devices is a challenging problem nowadays due to variations in style, color, background complexity, influence of luminance etc. Automatic detection of text in a natural scene image is useful to blind and foreigners with language barrier. Furthermore, it also has potential applications in robotics, image retrieval and intelligent transport systems.

A survey work of existing methods for detection, localization and extraction of texts embedded in images of natural scenes can be found in [1]. Two broad categories of available methods are connected component (CC) based and texture based algorithms. Earlier, Wu et al. [2] proposed a texture segmentation method to generate candidate text regions. A set of feature components is computed for each pixel and these are clustered using K-means algorithm. More recently, Jung et. al. [3] employed a multi-layer perceptron classifier to discriminate between

S.O. Kuznetsov et al. (Eds.): PReMI 2011, LNCS 6744, pp. 446–451, 2011.

text and non-text pixels. Considering Bangla/Devnagari script, Bhattacharya et. al. [4] proposed a scheme based on analysis of CCs for extraction of Devnagari and Bangla texts from camera captured outdoor images. Also a few criteria for robust filtering of text components have been proposed. Bangla and Devnagari are two most popular Indian scripts used by more than 200 and 500 million people respectively in the Indian subcontinent. An unique and common characteristic of these two scripts is the existence of certain headlines that act as an interlink among symbols of a word. We here, take an interest into color images embedding text. We first extract a number of color and shape based features from an input color image. The fuzzy c-means clustering is applied on these features, for segmentation (Sect. 2). After color image segmentation, headlines are detected from each CCs using morphology on the skeleton image (Sect. 3). Afterwards, the CCs attached with these headlines are separated. These separated components may contain both text and non-text. We apply certain shape and position based purification in order to distinguish between text and non text (Sect. 4). However some text portions, that are not connected with headline, will not be separated by the above procedure. To cope this, we increase the area of the bounding box of each CC to a specified limit. The portions that lie inside this modified bounding box are now studied against the previous criteria to obtain text (Sect. 5).

Concerning the dataset, to the best known, no benchmark dataset of outdoor images consisting Bangla and/or Devnagari is available. The present study is based on a set of 100 outdoor images. Initial tests (Sect. 6) show well separation of texts from the images.

2 Color Image Segmentation

Color image segmentation is our first step of text extraction. The fuzzy c-means algorithm is used for color image segmentation. Before applying c-means, we extract some features from the normalized RGB image. Let us consider a pixel p_i of the image. Then p_i can be described by the tuple (r_i, g_i, b_i) i.e. the normalized R, G and B values. Besides, these three color values, we take another shape based feature of p_i. Inside an object (text here), intensity of pixels are assumed to be homogenous. We consider an $n \times n$ window surrounding p_i. Let β_i be the number of pixels having same intensity as p_i inside the $n \times n$ window. Then β_i is our another feature. Combining, the feature vector ($\mathbf{f}_i$) corresponding to the pixel p_i is: $\mathbf{f}_i = (r_i,\ g_i,\ b_i,\ \beta_i)$. These features are sent to the fuzzy c-means clustering procedure.

3 Morphology Based Headline Detection

Most of the Bangla and/or Devnagari characters are connected by the headline. Hence, in order to extract the Bangla and/or Devnagari text portions we should detect the headline that joins them. Here, we apply mathematical morphology operation to obtain the lines inside the image. Further, applying some filtering we are able to detect the headline. The procedure works as follows. At first sufficiently small and large CCs are removed. For each remaining CCs, the skeleton image is constructed by morphological operation. Let us denote the skeleton image by A. With the skeleton image, we perform morphological opening operation to extract straight lines. It is evident that opening of an image A with a linear structuring element B can effectively identify the horizontal line segments present in a CC. However, a suitable choice of the length of this structuring element is crucial for processing at the latter stages and we empirically set this length as 21 pixels for the present dataset. The effect of skeletonisation removes pixels on the boundaries of the image but does not allow components to break apart. After skeletonisation, only the required pixels remain and thus a less number of lines detected. However, if no such lines are detected, we need to apply morphology on the original CC. Also, we may encounter a component small enough that we could not find any horizontal line. Such components remain as it is.

Now we have a set of lines over the image. We next apply some simple criteria to separate out possible headlines. Let a detected line be denoted by L_i. Let H_u and H_l be the heights of the portions of the corresponding CCs lie at the upper and at the lower half of the line L_i. Bangla characters mostly lie at the lower portion of the headline, only a small portion may reside at the upper of the headline. So, for a headline L_i we should have $H_u < H_l$. Thus at the next step we sort out all the lines L_i for which $H_u < H_l$. These lines mostly represent the headlines.

4 Headline Attached Text Portion Separation

After finding the headlines, the components attached with these headlines are separated. These components may include text as well as non-text. However, applying some text specific conditions we are able to separate out text components. These conditions are described below. All the concerned components are subjected to these conditions in the same sequence as given. One important note here, is, the thresholds specified in one condition are found after performing the previous conditions. Thresholds may differ if the sequence is altered.

1. *Removal of boundary attached connected components*: Generally, text like patterns are not attached with boundary of the image. So, we first remove all boundary attached CCs using morphological reconstruction.
2. *Elongatedness ratio*: For elongatedness ratio (ER) we use the measure designed by Roy et. al. [5]. Empirically it is found that a component with ER value greater than 5 is a text part.
3. *Number of holes*: Usually, text like patterns contain less number of holes than non-text patterns. Using the Euler number, we calculate the number of holes inside a component. Found empirically, a text component has less than 9 holes.
4. *Aspect ratio*: The aspect ratio of a non-text component is either very small or very large. We found by experiments, that the aspect ratio of a text component becomes less than 0.3 or greater than 2.0.

5. *Object to background pixels ratio* (r): This measure is computed inside the component bounding box. Due to the elongated nature of Bangla and Devnagari texts, only a few object pixels fall inside bounding box. On the other hand, elongated non-texts are usually straight lines, so, contribute enough object pixels. We observe $0.3 \leq r \leq 3$ could identify text components.

After the above procedure, all the headline attached text components are separated.

5 Identification of Headline Unattached Text

As discussed above, we successfully separate headline attached text portions. However, in Bangla/Devnagari some text components do not meet the headline. Now we consider such components. These components, though not connected, must be close to one/more of already detected text components. Then, if we increase area of the component bounding box enough, the possible text components may lie inside it. With this view, we increase the width of the bounding box by its height and the height by an empirical threshold. Now, the components inside this modified bounding box are subjected to the text identification conditions (Sect. 5).

6 Results and Discussion

In this section we present the results after applying our algorithms to a set of outdoor images from our dataset. Our dataset consists of 100 test images captured by a DSC-W320/P SONY digital still camera (14.1MP). The images contain Bangla and/or Devnagari texts.

Let us first consider the "X-Ray" image (Fig. 1(a)) as an example. Here, besides the arrow and thin boundaries, the text components are in white. The segmented image is shown in Fig. 1(b). Fig. 1(c) presents the headlines detected by morphology. The components that are attached with these headlines are shown in Fig. 1(d). Compering this with Fig. 1(a), we notice, all but three text components are present. These three components are unattached to the headline. Now, we perform the elimination of boundary attached components. As a result (Fig. 1(e)), the big blob at the top of the image is now eliminated. However, the non-text portion present at the bottom of the image is not boundary connected, so remain unchanged. Here, we can filter out this component successfully after testing the elongatedness ratio (Fig. 1(f)). However, the background component present here is elongated enough, so satisfy this criteria. We observe a large number of holes inside this component. Hence, after putting a condition on number of holes we may remove this component as well (Fig. 1(g)). The next two operations i.e. the aspect ratio and the ratio of object and background pixels, do not affect the results. Afterwards, we perform the identification of text components unattached to headlines. Fig. 1(h) gives the result.

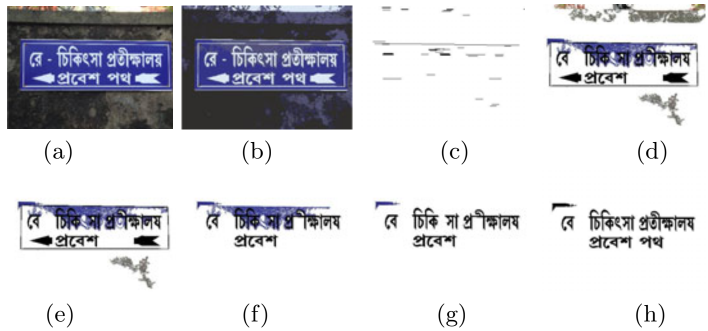

Fig. 1. (a) Input "X-Ray" image, (b) after segmentation, (c) after detecting headlines, (d) all CCs attached with headlines, (e) after removal of boundary attached CCs, (f) after performing elongatedness ratio, (g) after filtering using number of holes, and (h) after retrieving headline unattached text

Fig. 2. (a), (b), (c), (d), (e) Sample images and (f), (g), (h), (i), (j) the corresponding segmented text portions

We may notice the components absent in Fig. 1(d) are now identified successfully. In Fig. 2, we present some more results on images from our dataset. As, Fig. 2 suggests, the skew and perspective issue can't affect our procedure to some extent. However, in heavily skewed images the algorithm may fail due to wrong detection of headlines. Also note the successful results of "Salbani" image (Fig. 2(e)) having poor lighting condition.

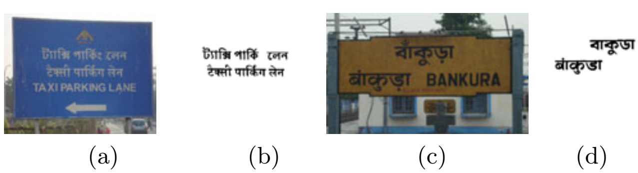

Fig. 3. Results on images showing degraded performance. (a), (c) The original images and (b), (d) the text portions.

Table 1. The precision and recall computation for our dataset

No. of Images	Status	Total text	Correctly recognized text	Wrongly recognized text	Precision	Recall
62	Perfectly extracted	307	307	27	91.91%	100.00%
32	Partially/extracted with some non text	147	91	29	75.83%	61.90%
06	Poor performance	27	09	31	22.50%	33.33%
100		481	407	87	82.38%	84.61%

The images in Fig. 3 show poor performance. The two images have headline unattached text portions that are not restored. The precision and recall values of our algorithm obtained on the basis of the present set of 100 images are respectively 82.38% and 84.61%. Detailed results are presented in Table 1.

7 Conclusion

This article provides a methodology that aids automatic extraction of visual text entities embedded in complex outdoor images. The proposed method is not very sensitive to image color, text font, skewness and perspective effects. Moreover it analyzes only headline attached CC adding a few CC which are closed to the headline. So, it becomes computationally efficient for real applications. This method can be extended to scanned documents also. The results shown are significant based on a laboratory made dataset. In future, we shall study the use of machine learning tools to improve the performance.

References

1. Liang, J., Doermann, D., Li, H.: Camera based analysis of text and documents: a survey. Int. Journ. on Doc. Anal. and Recog (IJDAR) 7, 84–104 (2005)
2. Wu, V., Manmatha, R., Riseman, E.M.: Textfinder: An automatic system to detect and recognize text in images. IEEE Trans. on Pattern Analysis and Machine Intelligence 21, 1224–1229 (1999)
3. Jung, K., Kim, I.K., Kurata, T., Kourogi, M., Han, H.J.: Text scanner with text detection technology on image sequences. In: Proc. of Int. Conf. on Pattern Recognition, vol. 3, pp. 473–476 (2002)
4. Bhattacharya, U., Parui, S.K., Mondal, S.: Devanagari and bangla text extraction from natural scene images. In: Proc. of the Int. Conf. on Document Analysis and Recognition, pp. 171–175 (2009)
5. Roy, A., Parui, S.K., Paul, A., Roy, U.: A color based image segmentation and its application to text segmentation. In: Proc. of Ind. Conf. on Computer Vision, Graphics & Image Processing, pp. 313–319 (2008)