Text Localization in Historical Document Images with Local Binary Patterns and Variance Models

Tapan Kumar Bhowmik[1] and Manika Kar[2]

[1] LITIS EA-4108, Université de Rouen, France
[2] Departamento de Engenharia Informática, Universidade do Porto, Portugal
tapan-kumar.bhowmik@univ-rouen.fr, manika.kar@fe.up.pt

Abstract. In this paper, we explore the utility of Local Binary Pattern (LBP) descriptors and variance measure towards the development of efficient techniques in order to segment a large collection of historical machine printed document pages. The result of segmentation will help us to organize the document pages in a structural format, which is useful in many applications like historical document access. In our experiments, three basic reference models namely background, text and image models are used to segment various non-text information together with the text. The method is tested on an archive of Portuguese historical documents and shows promising results.

1 Introduction

In recent years, the document analysis and recognition community has shown growing interests in processing of historical documents. These old documents often have historical and cultural significance and the aim is to scan them and create digital libraries [2], thereby offering continuous electronic access to this important part of the cultural heritage. To build such digital libraries, the first challenging task is to segment and classify the digitized documents into regions of text and image so that we can represent and store the scanned documents in a structural format. Representing documents in this way facilitate easy access to many non-textual information together with the text. During digitization process, in case of historical machine printed documents, OCR technology is used to convert text images into ASCII format for easy storage and retrieval. But, current OCR technology is largely restricted to find text printed against clean background, and cannot handle properly text printed against shaded or texture background, and/or embedded in images. However, more sophisticated text reading systems usually employ layout analysis schemes (page segmentation) for identifying text regions before applying OCR. But most of them [10, 8, 4, 5, 12] require fair background and the techniques are based on connected component analysis, which works on binarized version of input images. On the other hand, historical documents often suffer from several degradations and as a result it is hard to get clean binarized version of an input image. Therefore, we need a page segmentation scheme which works on natural background of the images without involving binarization.

In this paper, we describe such a text localization scheme based on multi-resolution analysis that we build as the first step of page layout system for historical documents. We assume that a document page has different texture regions due to different backgrounds, texts and images. So, we consider the page segmentation problem as a texture segmentation problem where segmentation process will partition the image into regions based on their texture [6, 3, 11, 7]. We explore the utility of various Local Binary Patterns (LBP), one of the most popular local image descriptor for texture analysis and use in our text localization scheme. In addition, variance measure is also used to localize the text. The proposed scheme is evaluated on an archive of Portuguese historical documents [1] and gives promising results.

2   Local Binary Pattern

Local Binary Pattern (LBP) was introduced by Ojala et. al. [9] for describing the local texture of gray scale images. The LBP operator describes each pixel (c) in an image with a certain binary pattern by calculating the difference of the gray values from a central pixel (g_c) around its circular neighborhood (with radius $R > 0$). If the difference of gray values between a neighboring pixel and the central pixel is greater than or equal to zero, the value is set to one, otherwise is set to zero (see in Figure 1(a)). Formally, the LBP operator is defined as follows:

$$LBP_{P,R}(c) = \{s(g_p - g_c)\}_{p=0}^{P-1} \simeq \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \text{ where}$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and $g_p(p = 0, 1, \ldots, P - 1)$ correspond to the gray values of P equally spaced pixels on a circle of radius R, which forms a circularly symmetric neighbor set. If the coordinates of g_c are assumed to be $(0,0)$, then the coordinates of g_p are calculated by $(Rsin(\frac{2\pi p}{P}), Rcos(\frac{2\pi p}{P}))$. Figure 1(b) illustrates circularly symmetric neighbor sets for various (P, R). The gray values of neighbors which do not fall exactly in the center of pixels are estimated by interpolation. Since, the difference of intensity is considered only while forming the binary pattern, so by definition the description is invariant against any monotonic transformation of the gray scale.

2.1   Rotation Invariance LBP

The basic $LBP_{P,R}$ operator produces 2^P different binary patterns that can be formed by the P pixels in the neighbor set. For example, $LBP_{8,1}$ operator can produce 256 unique binary patterns, which are not rotationally invariant. These are made rotation invariant by selecting the smallest value of $P - 1$ bitwise shift operations on the binary pattern. Formally, the operator is defined as:
$LBP_{P,R}^{ri}(c) = min\{ROR(LBP_{P,R}(c), i) | i = 0, 1, \ldots, P - 1\}$, where $ROR(., i)$
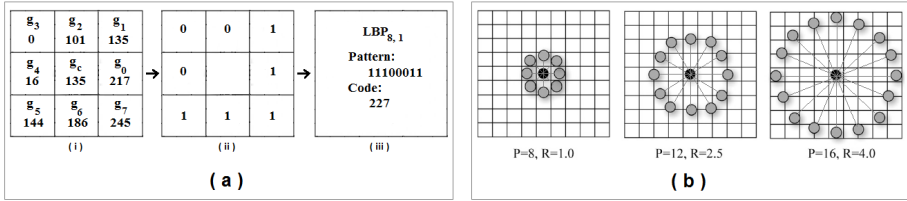
Fig. 1. (a) Example for calculating the LBP code. (b) Three circularly symmetric neighborhood sets (gray) around a central pixel (black) for different (P, R). If a point does not fall exactly on a pixel grid, the value is interpolated.

performs a circular bit-wise right shift on the P-bit binary pattern (say, $\{s(g_p - g_c)\}_{p=0}^{P-1}$) i times. For example, rotation invariant of $LBP_{8,1}(c) = 227$ in Figure 1(a) is $LBP_{8,1}^{ri}(c) = min\{227, 241, 248, 124, 62, 31, 143, 199\} = 31$. It is also be noted that there are 36 unique rotation invariant local binary patterns that can occur in the case of $LBP_{8,R}^{ri}(c)$ [9].

2.2   Uniform and Non-uniform LBP

From the practical experience, it was observed by Ojala et. al. [9] that certain binary patterns are fundamental properties of texture, providing the vast majority of patterns, sometimes over 90%, among all 3×3 patterns present in the observed textures. From this intuition, uniform and non-uniform patterns are defined. A pattern is considered as a uniform, symbolically called $LBP_{P,R}^{riu2}(c)$, if the number of transactions in the sequence between 0 and 1 are less than or equal to two. For example, 00000000_2 and 00000010_2 are the uniform patterns, as they have exactly zero and two 0/1 transitions respectively. Similarly, 10000010_2 is the non-uniform pattern because it has three 0/1 transitions. Non uniform patterns are considered the patterns that contain the main part of the noise of the images. Notice that a region with no transition is either a background or a flat region of the image. It is also be noted that, when $P = 8$, there exists only 9 uniform patterns out of 36 unique rotation invariant patterns [9].

2.3   Representing Image with LBP

For describing an image with LBP, each and every pixel is examined and found the distribution of the binary patterns for the whole image. It is fact that the different texture images have different distributions of binary patterns. This is done by calculating the histogram of binary patterns. The number of bins in the histogram depends on the choice of the operator. For example, the number of bins for $LBP_{8,R}(c)$ and $LBP_{8,R}^{ri}(c)$ are 256 and 36 respectively. It is observed that uniform patterns are present in distributions a lot more frequent than non-uniform patterns. Giving more weight to uniform patterns provide better discrimination. This is done by grouping all non-uniform patterns into single bin
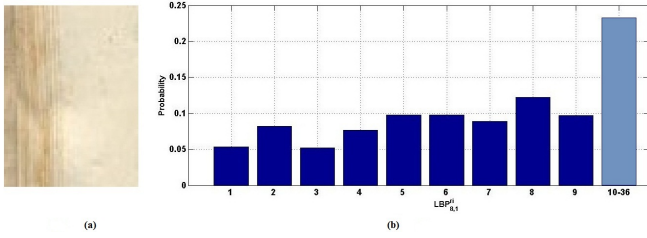
Fig. 2. Representation of an image with LBP. (a) image and (b) its corresponding $LBP_{8,1}^{riu2}$ histogram.

of the distribution. For example, in case of $LBP_{P,R}^{riu2}(c)$ operator the number of uniform pattern is 9 and non-uniform is $36 - 9 = 28$. But the histogram with number of bins 10, each uniform pattern is assigned to separate single bin and all the non-uniform patterns are assigned to another single bin (see in Figure 2).

3 Variance Measure

The variance is estimated in measure of standard deviation in local region of an image, namely window/block. Same LBP operators (described in previous section) are used to get a center pixel and its neighbor pixels (see in Figure 1(a)(i)). For a pixel (c) in a block image and a given operator $LBP_{P,R}(c)$, the standard deviation is measured by $\sigma(c) = [\frac{1}{P}\{(g_c - \mu(c))^2 + \sum_{p=0}^{P-1}(g_p - \mu(c))^2\}]^{\frac{1}{2}}$, where mean $\mu(c) = \frac{1}{P}[g_c + \sum_{p=0}^{P-1} g_p]$. If there is N number of center pixels (c) in a block image (w), the average variance measure is calculated as: $\sigma_{avg}(w) = \frac{1}{N}\sum_{c=0}^{N}\sigma(c)$.

4 Proposed Text Localization Approach

The input image is first divided into logical grids with fixed size disjoint window. For each window, LBP histogram feature is extracted. While building histogram feature, three LBP-operators $LBP_{P,R}^{ri2}(c), R = 1, 2, 3$ are used simultaneously due to preserving the multi-resolution property as well. For R equal to 1, 2 and 3, the ideal value of P is 8, 16 and 24 respectively. During the binary pattern calculation, only $P = 8$ equidistant points are considered. The remaining points (in case of $P = 16, 24$) help to find the gray values of 8 points. For example, when $P = 16$, the gray value (g_{p_i}) of a point p_i is considered the median gray value of three (2R - 1) consecutive points say, $p_{i-1}p_ip_{i+1}$. Similarly, in case of $P = 24$ five points are considered to choose the gray value of a point.

4.1   Building Reference Models

It has been observed that most of the historical documents composed with three basic regions: background (white or black), text and image. As an initial target, for separating text region from background and image, three reference models are considered from each type of region: i) background model ii) text model and iii) image model. The background models are built from the white and black portions of the images, image models are built from the image portions while text models are learnt from the text portions of various collection of complex document images. For each of the three reference models using LBP, normalized values of each bin in LBP-hisotgram are used as the model parameters. The comparison of such a set of three reference models (image, text and background) are shown in the Figure 3. On the other hand, for each of the three reference models using variance, mean (μ) and standard deviation (σ) are the model parameters.
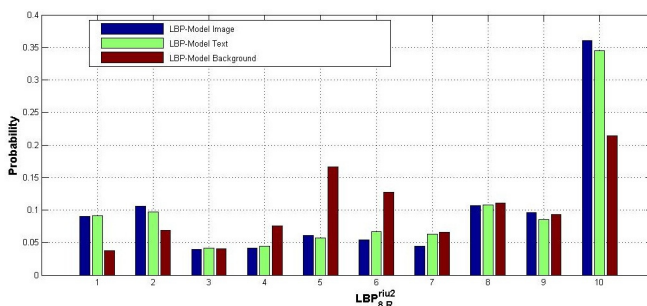

Fig. 3. Comparison of three reference LBP models (Image, Text and Background)

4.2   Measuring Model Response

Let us assume that $S_{w_{(i,j)}}$ is the sample by means of normalized LBP-histogram estimated from a certain window image ($w_{(i,j)}$). The model response for the sample S is defined as $L(S, M) = \sum_{b=1}^{B} S_b \log M_b$. It is also be noted that a block (a sample S) is assigned to the class of the model M that gives the maximum response. Similarly, the response of variance model is calculated from the normal distribution associated with the corresponding model parameters say mean (μ) and variance (σ^2).

4.3   Block Re-labeling with Hierarchical Multiresolution Analysis

Initially, all the blocks are assigned to a certain label say, either it is background or text or image based on the model response. Let us assume that if a block is classified as "image", it is marked with number 0, if it is "text" then is marked

with number 1 and for "background" is marked with number 2. In block re-labeling process we re-assign the blocks further on the basis of the following criteria: i) If a block and its 8-neighbor blocks are in same label, then keep the label as it is. ii) If a block and its 8-neighbor blocks are in different labels but all the 8-neighbor blocks are in same label, then reassign the block with the neighbor's block label. iii) If a block and its 8-neighbor blocks are in different labels but all the 8-neighbor blocks are also not in same label, then all the neighbor's blocks are merged with its central block and resize it into previous block size, then extract the feature again from that block and examine the model response. If it is same as previous one, then the block label keep as it is, otherwise majority voting scheme is applied with the neighboring labels and change the block label accordingly.

5   Experiments

First, we divide each document page logically into grids of a fixed window size. The size of the window is chosen based on the resolution of the image and it is calculated by the equation $w = \lfloor 3\lambda\sqrt{d} \rfloor$, where d is the resolution of the image measures in dpi and λ is a constant term. From our experiment, it is found that the localization scheme works well when $\lambda = 1$. Separate experiments have been performed using LBP and variance model. Some initial segmentation results are shown in the Figures 4 and 5. In the figures, the background, text and image regions are marked with blue, green and red windows respectively. Though from Figure 4 it is shown that the model $LBP_{8,R}$ gives slightly better result than the other two models $LBP_{8,R}^{ri}$ and $LBP_{8,R}^{riu2}$, but in many cases it is found that all the three LBP models produce almost similar results. However, from the experiments, we observe that the variance models are very effective to separate text and image regions from background but sometimes it fails to separate image region from the text (see in Figure 5(d)). On the other hand, when image regions (specific to document corpus, like as historical documents) have well defined number of texture patterns, LBP is very useful to separate image region from the text, even though both the text and image regions have similar variance. Such cases are shown in Figure 5. It is a fact that the performance of segmentation depends on the size of the window. So fixed-size window approach sometimes cannot handle documents that have text lines with capital letters or large fonts (such as heading, title) and it is likely to be classified as background region. To overcome this situation, we re-assign the label further with multi-resolution analysis (described in section 4.3). In addition, the re-labeling process helps to remove many erroneous labeling as well. Few such interesting results of the final segmentation (after re-assigning the label) are shown in the Figure 6(a) and Figure 6(b). We have evaluated our algorithm on 20 images ($2 * 20 = 40$ pages) from the archive of Portuguese historical documents [1] and overall performance of LBP and variance models are given in Table 1.
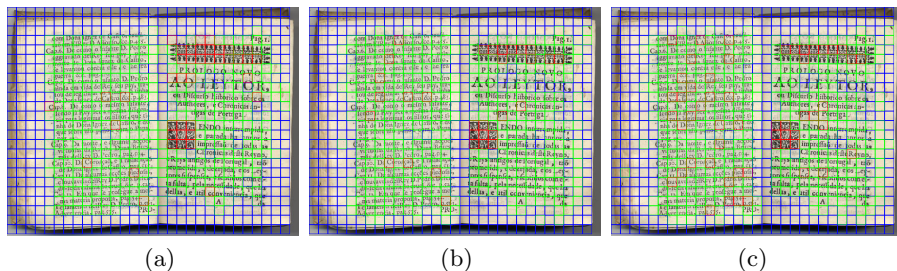
Fig. 4. Page segmentation of Portuguese document with LBP models. The outputs (a), (b) and (c) are produced by the three LBP operators $LBP_{8,R}$, $LBP_{8,R}^{ri}$ and $LBP_{8,R}^{riu2}$ respectively. The regions image, text and background are marked with red, green and blue colors respectively.
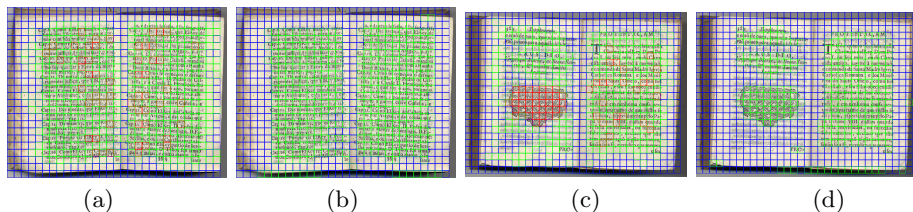

Fig. 5. Page segmentation of Portuguese document with LBP models. The outputs (a), (b) and (c) are produced by the three LBP operators $LBP_{8,R}^{riu2}$ and var respectively. The regions image, text and background are marked with red, green and blue colors respectively.
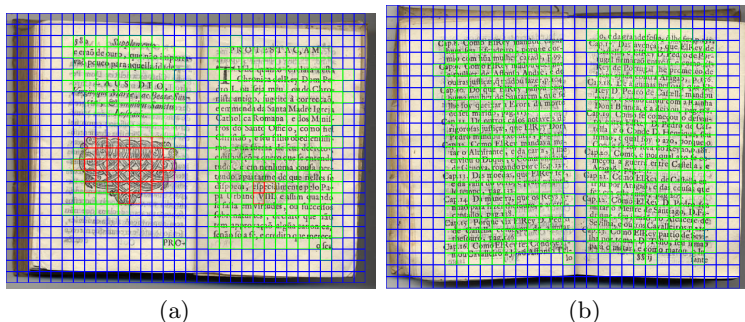

Fig. 6. Page segmentation of Portuguese document with (a) LBP models and (b) Variance models. Blue color indicates the background region and green color represent the text region.

6   Conclusion and Future Scope

In this paper, we explain a text localization scheme for historical documents with local binary pattern and variance model. We also introduce reference models concept in the intermediate stage to explore more difficulties for this kind of

Table 1. Text localization results on the archive of historical documents

No. of pages	No. of text zones	LBP		Variance	
		Recall (%)	Precision (%)	Recall (%)	Precision (%)
40	74	97.78	88.60	76.72	93.61

problems. As an initial experiment, three reference models background, text and image models are used from each category to identify the text region from background and image. The novelty of this framework is that we can add many more reference models depending upon the complexity of the document's archive. It is not necessary that only single text model can be used, multiple text models can be created from different classes of text textures. However, the response values from different models can be used to generate another level of features for further classification. In this stage, many sophisticated classification techniques such as SVM, MLP can be applied to separate text regions from the documents more accurately.

References

1. Biblioteca Nacional De Portugal, `http://purl.pt/index/geral/PT/index.html`
2. Baird, H.: Digital libraries and document image analysis. In: Proc. of the 7th ICDAR, pp. 2–14 (2003)
3. Etemad, K., Doermann, D., Chellappa, R.: Multiscale segmentation of unstructured document pages using soft decision integration. Trans. on. IEEE 19(1), 92–96 (1997)
4. Fletcher, L.A., Kasturi, R.: A robust algorithm for text string separation from mixed text/graphics images. Trans. on. IEEE 10(6), 910–918 (1988)
5. Gorman, L.O.: The document spectrum for page layout analysis. Trans. on. IEEE 15(11), 1162–1173 (1993)
6. Jain, A., Bhattacharjee, S.: Text segmentation using gabor filters for automatic document processing. Machine Vision Appl. 5, 169–184 (1992)
7. Kim, K.I., Jung, K., Kim, J.H.: Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. Trans. on. IEEE 25(12), 1631–1639 (2003)
8. Nagy, G., Seth, S.C., Stoddard, S.D.: Document analysis with an expert system. In: Pattern Recognition in Practice II, pp. 149–155. Elsevier Science, New York (1986)
9. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classi cation with local binary patterns. Trans. on. IEEE 24(7), 971–987 (2002)
10. Wong, K.Y., Casey, R.G., Wahl, F.M.: Document analysis system. IBM J. Res. Development 6, 456–642 (1982)
11. Wu, V., Manmatha, R., Riseman, E.M.: Textfinder: An automatic system to detect and recognize text in images. Trans. on. IEEE 21(11), 1224–1228 (1999)
12. Zheng, Y., Li, H., Doermann, D.: Machine printed text and identification in noisy document images. Trans. on. IEEE 26(3), 337–353 (2004)