2009 10th International Conference on Document Analysis and Recognition

A Variational Bayes Method for Handwritten Text Line Segmentation

Fei Yin，Cheng-Lin Liu
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, P.R. China
E-Mail: {fyin, liucl}@nlpr.ia.ac.cn

Abstract

Text line segmentation in unconstrained handwritten documents remains a challenge because handwritten text lines are multi-skewed and not obviously separated. This paper presents a new approach based on the variational Bayes (VB) framework for text line segmentation. Viewing the document image as a mixture density model, with each text line approximated by a Gaussian component, the VB method can automatically determine the number of components. We extend the VB method such that it can both eliminate and split components and control the orientation of text line lines. Experiments on Chinese handwritten documents demonstrated the effectiveness of the approach.

1. Introduction

Text line segmentation is one of important pre-processing tasks for document image analysis because it provides crucial information for word/character segmentation and text recognition. Despite the enormous efforts in layout and text line segmentation of printed and handwritten documents [1-9], the segmentation of text lines in unconstrained handwritten documents remains a major challenge because the handwritten text lines may be curved, multi-skewed, and the space between lines is not obvious.

The existing methods of text line segmentation can be roughly grouped into two classes: top-down and bottom-up. Projection profile analysis is a representative top-down method. To overcome its limitation to parallel text lines, piecewise projection method has been proposed [3], which detects text line segments from the projection on strips of image and then connect into text lines. The water flow based method [4] flows water from left and right sides of the image to find text lines, and labels the un-wetted areas finally. Bottom-up methods group small units of image (pixels, connected components, etc.) into text lines and then text regions. The smearing-based method merges pixels into text lines according to horizontal run-lengths [5]. Some methods group connected components (or similar blocks) into text lines according to proximity and line continuity [2][6] or using the Hough transform [7][8]. These methods either need careful design of proximity measure or need sophisticated post-processing in the Hough space. The recent method of Li et al. detects the boundary of text lines using the level set method [9]. Most existing methods have limitations when applied to unconstrained handwritten documents because they more or less assume horizontal, straight, parallel and un-touched text lines.

In this paper, we propose a new text line segmentation approach based on mixture density estimation using the variational Bayes (VB) framework. Viewing the document image as a distribution of pixels, each text line can be modeled as bivariate Gaussian distribution and the document is a mixture of Gaussians. The VB method can automatically determine the number of components without additional overhead of model order selection. For processing document images, we have extended the VB method such that it can split components as well as eliminate redundant components and selectively control the orientation of text lines. The effectiveness of the proposed approach has been demonstrated in experiments on Chinese handwritten documents.

2. Rationale

A document image is a collection of black pixels, which can be approximated as a probability density distribution of pixels. Some previous works have used density models as auxiliary measures for text line segmentation [3][9]. A text line, which is approximately straight (if a text line is curved, it can be split into multiple straight ones), can be modeled as a Gaussian density distribution. Thus, the document image is a mixture of Gaussian densities, with each component corresponding to a text line. To estimate the density parameters and determine the number of

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.98

IEEE computer society

components is not trivial, however. The maximum likelihood EM algorithm can estimate the mixture density given the number of components. To select the number of components using a model selection criterion, such as minimum description length (MDL) or minimum message length (MML), the EM algorithm needs to perform repeatedly on many different numbers of components. On the other hand, the variational Bayes (VB) method can automatically determine the number of components, and thus, is appropriate for text line segmentation because the number of text lines is unknown a priori.

2.1. Mixture Density and Variational Bayes

A document image is formulated as set of observation bivariate vectors $X=\{x_{1,\ldots,}x_N\}$, in which every element x_n comprising the (x,y) coordinates of a black pixel. Each black pixel x_n is associated with a latent variable (K-dimensional vector) $z_n=(z_{n1},\ldots,z_{nK})^T$, with $z_{nk} \in \{0,1\}$ denoting the membership of pixel x_n to the k-th density component (text line). Under the mixture model, the density of a pixel is formulated as

$$p(x_n, z_n \mid \Theta) = \prod_{k=1}^{K} \pi_n^{z_{nk}} p(x_n \mid \theta_k)^{z_{nk}}, \qquad (1)$$

where K is the number of text lines, and $\pi_1,\ldots,\pi_k$ are the mixing probabilities satisfying $\pi_k \geq 0$ and $\sum_{k=1}^{K} \pi_k = 1$; θ_k is the set of parameters defining the the *k-th* component, and $\Theta = \{\pi_1,\ldots,\pi_K;\theta_1,\ldots,\theta_K\}$ is the complete set of parameters.

For document images, it is reasonable to approximate a text line use a Gaussian density function because the contour of an elliptical Gaussian density resembles the long-shaped outline of text lines. Thus, the likelihood of a document containing N black pixels is

$$p(X,Z \mid \Theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \mathcal{N}(x_n \mid \mu_k, \Lambda_k^{-1})^{z_{nk}}, \qquad (2)$$

where each component is a Gaussian density with mean μ_k and precision matrix Λ_k. $\Theta = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}\}$ is the set of parameters.

The variational Bayes (VB) framework [10] is an effective method to simultaneously estimate the density parameters and the number of components. We outline the VB method below, and more details can be found in [11].

In mixture density estimation by EM, the evaluation of the posterior distribution $p(Z|X)$ of the latent variables Z maybe infeasible either due to the high dimensionality of latent variables (as is the case for document images) or because the posterior probabilities are not analytically computable (e.g., in the case of continuous latent variables). The VB method is an approximation scheme that maximizes a lower bound of the model evidence $p(X)$, which is decomposed into

$$\ln p(X) = \ell(q) + KL(q \parallel p), \qquad (3)$$

where

$$\ell(q) = \int q(Z) \ln\left\{\frac{p(X,Z)}{q(Z)}\right\} dZ, \qquad (4)$$

$$KL(q \parallel p) = -\int q(Z) \ln\left\{\frac{p(Z \mid X)}{q(Z)}\right\} dZ. \qquad (5)$$

The low bound $\ell(q)$ is maximized with respect to the distribution $q(Z)$ approximating $p(Z|X)$. If we allow any possible choice for $q(Z)$, the maximum of the lower bound occurs when the KL divergence vanishes, which occurs when $q(Z)$ equals the posterior distribution $p(Z|X)$. Suppose $q(Z)$ can be factorized as $q(Z) = \prod q_i(Z_i)$ and $q_i(Z_i)$ is analytically tractable, $p(Z|X)$ can be easily obtained in iterative re-estimation.

In the VB estimation of Gaussian mixture (also called VBEM), the parameters $\boldsymbol{\pi}$ and $(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ have conjugate priors Dirichlet distribution and Gaussian-Wishart distribution, respectively. The variational distribution $q(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ is factorized into

$$q(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = q(\mathbf{Z}) q(\boldsymbol{\pi}) \prod_{k=1}^{K} q(\mu_k, \Lambda_k), \qquad (6)$$

and the factorized distributions are iteratively updated by re-estimation like EM, wherein a responsibility $E[z_{nk}] = r_{nk}$ of a data point to a Gaussian component is estimated. During iteration, some of the mixture components that have very small responsibility from the data points can be removed. Thus, the number of mixture components is automatically determined.

3. Text Line Segmentation Algorithm

The computation of variational Bayes (VB) method on a document image with a large number of black pixels is still appreciable. To save computation, we first reduce the resolution of image using the Gaussian pyramid technique. The original image in 300DPI is reduced to 1/16 of pixels in 75DPI. The low resolution image (gray scaled) in the Gaussian pyramid is converted to binary by retaining all the pixels of non-zero intensity as black pixels.

At the reduced resolution, the text lines and the boundary between them can be identified without obvious loss. After mixture density estimation by the

VB method, the number of components (text lines) is obtained, the density parameters are then remapped to the original image, and the black pixels of the original image are assigned to text lines. Last, a heuristic post-processing procedure is taken to adjust some text lines.

3.1. Text Line Detection Using VB

The VB based mixture density estimation performs on the reduced resolution (75DPI) image. The overall approach is depicted in Fig. 1. First, we initialize the mixture model by a heuristic segmentation technique, and then the initial estimate of number of text lines and the Gaussian parameters are fed into the VBEM algorithm. On convergence of VBEM, for better fitting the mixture density with an improper initial number of components, we extend the VB method such that it can split components during optimization. Splitting also help alleviate the local optimum of VBEM. The initialization and components splitting procedures are detailed as follows.

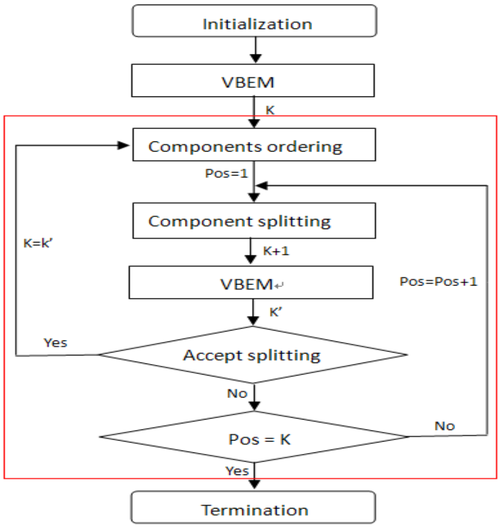
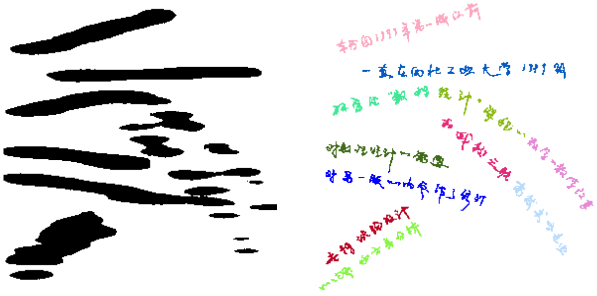


Fig. 1. Flow chart of text line detection using VB. The red box denotes the procedure of components splitting.

3.1.1. Initialization

To overcome the computation overhead of VBEM starting with a very large number of components, we use a heuristic technique similar to the one in [9] for initial estimate of number of text lines and density parameters. First, an anisotropic Gaussian kernel is used to blur the image (the window size is 120 pixels wide and 30 pixels high in our case), then the blurred image is binarized using the Otsu's algorithm. The number of connected components which is larger than an empirical threshold in the binarized image is taken as the initial number of Gaussian components, and the parameters of each Gaussian component are estimated from the black pixels of a connected component. Though this initialization procedure favors nearly horizontal text lines, an improper initial number of components on skewed text lines can be corrected by the following VB optimization. An example is shown in Fig. 2(a), where the skewed text lines are not properly segmented, but the VB method gives a proper final segmentation.



(a) (b)

Fig. 2. (a) Initial segmentation with 15 components; (b) Final result with 11 components by the VB method.

3.1.2. Aggressive Pruning VBEM

For density fitting on a document image, mostly the responsibility of Gaussian components does not shrink sufficiently because the text lines are not fitted perfectly by a mixture density model. Often, a redundant component responds to a few pixels which are unlikely to form a real text line. Therefore, we take a more aggressive pruning strategy, by which a component with the number of responding points (number of black pixels assigned to the component) is smaller than an empirical threshold (500 pixels in our case) is removed. As shown in Fig. 1, an initial estimate with 15 components in Fig. 2(a) is reduced to 11 components in Fig. 2(b) after VBEM optimization.

3.1.3. Dynamic Splitting

Since our initialization cannot guarantee that the initial number of components is larger than the actual number of text lines, we extend the VB method to enable components splitting. Inspired by the work of [12], we introduce a heuristic splitting mechanism, as shown in the red box of Fig. 1. Splitting starts from the K components after initialization and VBEM. The K components are ordered in preference of splitting. The top-ranked component (Pos=1) is selected to be split (replaced with two child components). The K+1 components are updated by VBEM to obtain $K'(\leq K+1)$ new components. The K' components are judged (criterion given later) to be accepted or not. If accepted, the K' components are re-ordered in preference and the top-ranked is selected to undergo splitting. Otherwise, the splitting is retracted and the next preferred component (Pos+1) is selected to undergo splitting. If Pos=K (all components have undergone splitting without being accepted), then the whole process terminates.

The preference of component splitting is measured by the second eigenvalue of the covariance. That is, a component having thick line shape is preferred to be split.

To split a selected component, the pixels assigned to this component are divided into two subsets according to the sign of the inner product with the principal eigenvector of the covariance. The two subsets of pixels form two child components to substitute the parent component.

After splitting and VBEM, the splitting is accepted if two conditions are met: (1) the component splitting increases the low bound $\ell(q)$ in Eq. (4) compared to that before splitting, and (2) the orientation of all the components are within the pre-defined range (this is to control the orientation of text lines, e.g., $-45^0 \sim +45^0$ for horizontal writing). Otherwise, the splitting is rejected and retracted.

Fig. 3 shows an example of splitting, where an initial estimate of four components in Fig. 3(a) is increased to 13 components in Fig. 3(b).

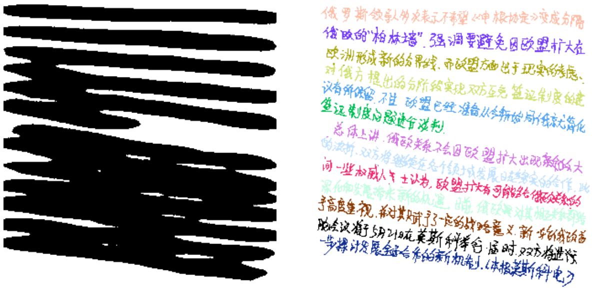


Fig. 3 (a) 4 components after initialization; (b) Final result with 13 components after splitting.

3.2. Post-Processing

On mixture density estimation in the low resolution image and remapping to the original image, most text lines can be grouped correctly. Some segmentation errors remain because the mixture density model does not fit the text lines completely. We take s heuristic post-processing procedure similar to [9] to adjust some text lines. Basically, if two text lines are nearly collinear and close to each other, they are merged into one line; if a text line has a large gap, it is split into two lines.

4. Experimental Results

We evaluated the performance of the proposed text line segmentation method on a Chinese handwritten documents database HIT-MW [13]. The database contains 853 pages written by more than 780 writers. It has 8,677 text lines and the pages were scanned at resolution of 300DPI. We have annotated the text lines and characters of all these document images using a ground-truthing tool.

We evaluate the performance by using measures similar to [14]: detection rate (DR, like recall rate), recognition accuracy (RA, like precision), missed detection rate (MDR) and false alarm rate (FAR).

We compare our approach with a piece-wise projection (PWP) method [1], which has shown superior performance in handwritten documents of English and Arabic scripts. On the 853 Chinese document images, the performance measures of our proposed method and the PWP method are shown in Table 1. We can see that the proposed method performs slightly better than the PWP in respect of DR and RA.

Table 1. Text line segmentation results.

	DR	RA	MDR	FAR
Proposed	0.9436	0.9477	0.0005	0.0004
PWP	0.9255	0.9310	0.0006	0.0007

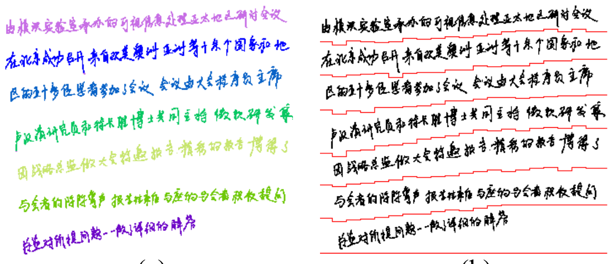


Fig. 4. Segmentation of horizontal text lines. (a) Proposed VB method; (b) Piecewise projection method.

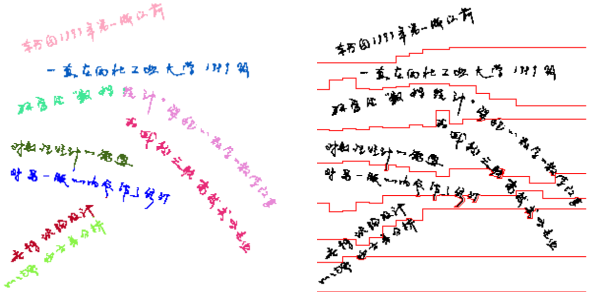
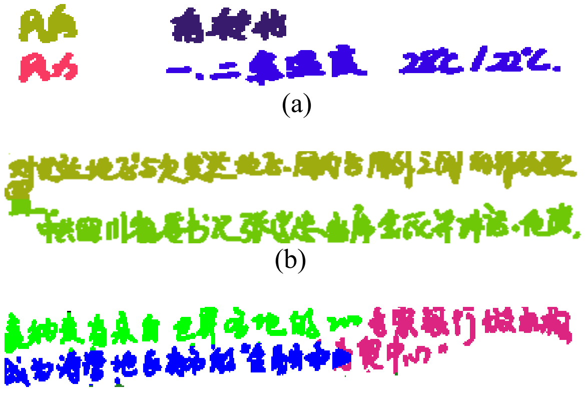


Fig. 5. Segmentation of multi-skewed text lines. (a) Proposed VB method; (b) Piecewise projection method.

Most of the actual documents consist of approximately horizontal and parallel text lines, so the proposed method and the PWP perform comparably well. An example is shown in Fig. 4. On the other hand, for documents with largely skewed text lines, the proposed method performs significantly well. An example is shown in Fig. 5.

The proposed method was implemented in C++ and experimented on a personal computer with CPU of Intel core2 3.0GHz. The processing time for a document image of 1700x1500 pixels is about 25 seconds.

Fig. 6 shows some segmentation errors left by the proposed VB method. In Fig. 6(a), the document was ground-truthed as two lines but was segmented into four lines. The document in Fig. 6(b) was ground-truthed as three lines (one of them has only one character) and was segmented into two lines. The document in Fig. 6(c) was ground-truthed as two lines but was segmented into three lines, one of which spans two correct lines and could not be corrected in post-processing.



(a)

(b)

(c)

Fig. 6 (a) a segmentation error; (b) a merge error; (c) an error including merge error and segmentation error.

5. Conclusion

We proposed a new method based on the variational Bayes (VB) framework for handwritten text line segmentation. By mixture density modeling with automatic determination of number of components, this method can detect multi-skewed, curved and slightly overlapping text lines. It can detect text lines of arbitrary orientation and also can flexibly control accepted orientations. Experimental results on handwritten Chinese documents demonstrate that the proposed method performs competitively with a superior piecewise projection method and has the advantage of handling largely skewed text lines. Future works will be to accelerate the VB method and combine the segmentation results of multiple algorithms.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC) under grant no.60775004 and no.60825301.

References

[1] L. O'Gorman, The document spectrum for page layout analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(11): 1162-1173, 1993.

[2] A. Simon, J.-C. Pret, A.P. Johnson, A fast algorithm for bottom-up document layout analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(3):, 273-277, 1997.

[3] M. Arivazhagan, H. Srinivasan, S. N. Srihari, A statistical approach to handwritten line segmentation, *Document Recognition and Retrieval XIV*, 2007, pp.6500T-1 to 6500T-11.

[4] S. Basu, C. Chaudhuri, M. Kundu, M. Nasipuri, D.K. Basu, Text line extraction from multi-skewed handwritten documents, *Pattern Recognition*, 40(6): 1825-1839, 2007.

[5] Z.X. Shi, V. Govindaraju, Line separation for complex document images using fuzzy runlength, *Proc. 1st Int'l Workshop on Document Image Analysis for Libraries,* 2004, pp.306-312.

[6] L. Likforman-Sulem, C. Faure, Extracting lines on handwritten document by perceptual grouping, In: *Advances in Handwriting and Drawing: A Multidisciplinary Approach*, 1994, pp.21-38.

[7] Y. Pu, Z. Shi, A natural learning algorithm based on Hough transform for text lines extraction in handwritten document, *Proc. 6th IWFHR*, 1998, pp.637-646.

[8] G. Louloudis, B. Gatos, I. Pratikakis, C. Halatsis, Text line detection in handwritten documents, *Pattern Recognition*, 41(12): 3758-3772, 2008.

[9] Y. Li, Y. Zheng, D. Doermann, S. Jaeger, Script-independent text line segmentation in freestyle handwritten document, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(8): 1313-1329, 2008.

[10] H. Attias, Inferring parameters and structure of latent variable model by variational Bayes, *Proc. 15th Conference on Uncertainty in Artificial Intelligence*, 1999, pp.21-30.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[12] Z. Ghahramani, M. J. Beal, Variational inference for Bayesian mixtures of factor analysis, In: *Advances in Neural Information Processing Systems 12*, 2000, pp.449-455.

[13] T. Su, T. Zhang, D. Guan, Corpus-based HIT-MW database for offline recognition of general-purpose Chinese handwritten text, *Int. J. Document Analysis and Recognition*, 10(1): 27-38, 2007.

[14] X.D. Zhou, D.H. Wang, C.-L. Liu, Grouping text lines in online handwritten Japanese documents by combining temporal and spatial information, *Proc. 8th IAPR Workshop on Document Analysis Systems*, 2008, pp. 61-68.