Darko BRODIĆ[1], Zoran N. MILIVOJEVIĆ[2], Dragan R. MILIVOJEVIĆ[3]

University of Belgrade, Technical Faculty in Bor (1), Technical College Niš (2), Institute for Mining and Metallurgy Bor (3)

Comparison of Two Goal-Oriented Methods for the Evaluation of the Text-Line Segmentation Algorithms

***Abstract**. Text line segmentation process represents the key step in the optical character recognition. Hence, the efficiency evaluation procedure for text line segmentation algorithms is the challenge. Text line segmentation process is established by the algorithms application to the text dataset. Furthermore, two goal-oriented methods for the evaluation of the text line segmentation results based on extended errors type and binary classification are explained. The paper presents the main points of the provided analyses and results discussion. The results confirm the superiority of the extended errors type over binary classification evaluation method.*

***Streszczenie.** Przedstawiono analizę algorytmów segmentacji linii tekstu. Analizowano dwie metody – analizy błędu i binarnej klasyfikacji. Wykazano przewagę pierwszej z tych metod. (**Porównanie dwóch algorytmów stosowanych do segmentacji linii tekstu**)*

Keywords: document image processing, text-line segmentation, image region analysis, image analysis, measurement.
Słowa kluczowe: przetwarzanie obrazu, segmentacja linii tekstu.

Introduction

Quality evaluation represents the most significant task in grading the algorithm for the document image processing [1]. Hence, the careful examination each of algorithms is of the primary importance. This process consists of detailed qualitative and quantitative analysis. In the experimental evaluation approach, the algorithm is applied to the initial data which represent a dataset. The obtained results are compared with the ideal results which represent the ground truth. Comparison is made with certain evaluation method. The objective is to correctly evaluate all characteristics of the examined algorithm. The whole process is based on two cornerstones: the appropriate test dataset and the efficient evaluation method.

Text line segmentation is one of the most important stages in document image processing [1]. It represents a labelling process, which consists in assigning the same label to spatially align units [2]. Any difficulties in this process stage will result in inaccurate segmented text lines. Furthermore, it will lead to the failure of the optical character recognition (OCR) [3].

Many algorithms have been developed for this task. They differ in the methodology, principle of work, efficiency, applicability, computer time consuming, etc. In their examination and quality evaluation the testing process is indispensable. The test is performed over the real data. The dataset is based mainly on the handwritten or printed text samples. However, they are consisting mostly on the English script [4-6] or extended by additional scripts [7-9]. For testing purposes, the algorithm is applied to the given datasets. The obtained results are classified according to the evaluation method.

The widespread technique is called the pixel-based method. It uses the comparison of the detected segmentation results with an already annotated ground truth [5,9,10]. Hence, if the ground truth line and the corresponding detected line share 90% or more pixels, the line is correctly detected [5]. This is an empirical guideline. Consequently, it is not valid in some circumstances [11-12].

Performance evaluation of the text line segmentation is a goal-oriented task. Few methodologies have been established on this premise [12-15]. This paper presents the evaluation methods which are based on the binary and the error type classification. At the end, the main parameters of two methods are compared as in [12].

The paper is organized in the following way. Section 2 gives a brief description of the experimental framework for the text line segmentation that contains the test evaluation procedure which involves the classification of text objects and text segmentation errors. Section 3 arranges these results according to binary and error type classification. Section 4 contains the comparative analyses and discusses both evaluation methods. Section 5 gives conclusions.

Experimental framework

Testing procedure represents the process of applying the algorithm to the dataset. Its assignment is to evaluate algorithm characteristics correctly. Hence, the text experiment should incorporate different classes of text types and scripts. Furthermore, many specific text line phenomena like touching lines, mixed lines, and indentation lines have to be included. Typically, dataset consists of the synthetic and handwritten text samples [12]. However, the major part of such phenomena is linked with the handwritten text. The following tests are exploited [12-15]: multi-line segmentation test of straight, waved, fractured text, and handwritten text.

Segmentation

The objective of the segmentation process is partitioning the image into regions. Let I be the set of indices from the area to be segmented, e.g. $I = [0,X)\times[0,Y)$ in a two-dimensional image with dimensions $X\times Y$ [16]. The segment is defined as a region which represents the partial set of I as $s_k \subset I$. Accordingly, $k = 1,\ldots,K$, where K is the total number of segments. Hence, the segmentation is partitioning of I into a set of regions that represent segments such that [16-17]:

- s_k is a connected region,
- $\bigcup_{k=1}^{K} s_k = I$,
- $s_k \cap s_l = \emptyset$, for all k and l, $k \neq l$.

Currently, there are two types of segments: segments which denote objects-to-detect and a segment consisting of the remaining area. Our segments of interest are only the first ones. In the following text, because of the nature of the algorithm the segments are named connected-components.

Classification of the text segmentation elements

During the test procedure, the algorithm is applied to the datasets which represents the text image samples. As the results, the text line segmentation is given. Obtained results correspond to the resultant text line extraction. These results are compared with text line ground truth and evaluated by selecting an evaluation method. This procedure is illustrated in Fig. 1.

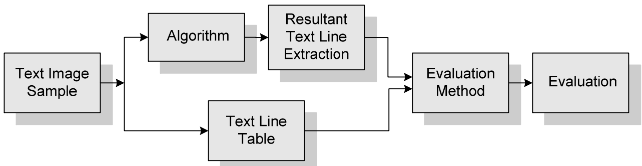


Fig. 1. Schematic test procedure for the algorithm evaluation.

From the test procedure the following elements can be defined [12]: initial connected components, resultant connected components and ground truth lines.

The number of connected components in the text image sample before the algorithm application represents the initial state and referred as initial connected components number CC_{init}. After the application of the algorithm to the text sample, the new layout of the connected components is established. The new number represents the resultant number of the connected components CC_{res}. However, the ideal number of text lines in the sample text represents its ground truth. It is called ground truth lines or connected components CC_{gt}. The algorithm efficiency is evaluated by comparing the ground truth with a resultant number of the connected components, i.e. CC_{gt} and CC_{res}. This procedure is illustrated in Fig. 2.

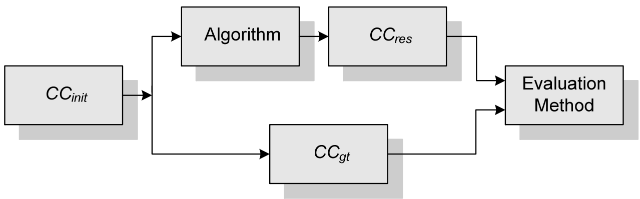


Fig. 2. Evaluation method elements.

Classification of the text segmentation errors

According to the relation between CC_{gt} and CC_{res} the result of segmentation is correct or incorrect. If certain CC_{res} cover just and only certain CC_{gt}, then this text line is segmented correctly. The total number of correctly segmented text lines in the text sample is marked as CC_{corr}. However, all others could be considered as error CC_{uncorr}. Segmentation errors belong to one of the following groups: over-segmentation, under-segmentation and mixed segmentation errors. Above classification is commonly referred as reduced classification scheme [17].

The circumstance where the text line is divided wrongly by the algorithm in two or more connected components is defined as over-segmentation error CC_{over}. In contrast, under-segmentation represents joined lines error. It corresponds to the situation where the sequence of two or more consecutive text lines is considered by the algorithm as a unique line. It is defined as under-segmentation error CC_{under}. Text lines including outlier words correspond to lines containing words that are incorrectly assigned to two adjacent lines. This circumstance is characterized as mixed segmentation error CC_{mix}. The sum of the all above mentioned errors give CC_{uncorr}. This classification is given in Fig. 3.

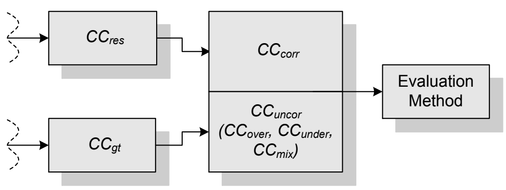


Fig. 3. Classification according to the text segmentation errors.

Evaluation methods

The evaluation method is performed according to the testing results. In this paper two evaluation methods will be presented: method based on errors type and method based on binary classification.

Evaluation method based on error type

The evaluate method based on error type is established completely on the text segmentation errors elements. Hence, according to the error elements CC_{corr}, CC_{over}, CC_{under}, and CC_{mix}, the following measures have been introduced [12]: segmentation line hit rate (*SLHR*), over-segmentation line hit rate (*OSLHR*), under-segmentation line hit rate (*USLHR*), mixed line hit rate (*MLHR*), and segmentation root mean square error ($RMSE_{seg}$).

SLHR represents the hit rate of the correctly segmented text lines. The hit rate is given as 1 - |*RE*|, where RE represents the relative error. Hence, *SLHR* is defined as:

$$SLHR = 1 - \left|RE_{corr}\right| = 1 - \left|\frac{CC_{gt} - CC_{corr}}{CC_{gt}}\right|. \tag{1}$$

All other measures are connected with the evaluation of the text line segmentation errors. Hence, they take into account the over, under and mix segmentation errors.

An increased number of the connected components per text line lead to the over-segmentation error. It is a consequence of the algorithm inability of merging all connected components from one text line into one connected component. *OSHLR* represents hit rate of the over-segmented text lines. It could be expressed as:

$$OSLHR = 1 - \left|RE_{over}\right| = 1 - \left|\frac{CC_{gt} - CC_{over}}{CC_{gt}}\right|. \tag{2}$$

If two or more consecutive text lines are considered as a unique one, then under-segmentation error is emerging. This process leads to a smaller number of connected components than the number of text lines. *USHLR* represents hit rate of the under-segmented text lines. It is:

$$USLHR = 1 - \left|RE_{under}\right| = 1 - \left|\frac{CC_{gt} - CC_{under}}{CC_{gt}}\right|. \tag{3}$$

The process of mutual injected connected components from different text lines leads to mixed segmentation error. This circumstance is similar to the over-segmentation one. However, it contains the residual connected components in the text line. *MLHR* represents hit rate of the mixed segmented text lines. It is:

$$MSLHR = 1 - \left|RE_{mix}\right| = 1 - \left|\frac{CC_{gt} - CC_{mix}}{CC_{gt}}\right|. \tag{4}$$

Furthermore, the number of resultant connected components and ground truth line per each line is compared. Its variance is given by $RMSE_{seg}$ as [15]:

$$RMSE_{seg} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(CC_{gt,i} - CC_{res,i}\right)^2}, \tag{5}$$

where N is the total number of lines in the text sample, $CC_{gt,i}$ is the number of ground truth lines in the text line i (equal to one per each line), and $CC_{res,i}$ is the number of resultant connected components in the text line i. This measure has been introduced for the fine evaluation of the algorithm primarily linked to the over-segmentation errors. However, it calculates residual elements in mixed and under segmentation errors, too. It incorporates mainly different phenomena like over, under and mixed segmentation text elements. As the extension to the given method, the measures $RMSE_{over}$, $RMSE_{under}$ and $RMSE_{mix}$ are proposed. $RMSE_{over}$ is defined as:

$$RMSE_{over} = \sqrt{\frac{1}{P}\sum_{j=1}^{P}\left(CC_{gt,j} - CC_{res,j}\right)^2}, \quad (6)$$

where P is the total number of over segmented lines in the text sample, $CC_{gt,j}$ is the number of ground truth lines in the over segmented text line j (equal to one per each line), and $CC_{res,j}$ is the number of resultant connected components in the over segmented text line j. $RMSE_{under}$ is defined as:

$$RMSE_{under} = \sqrt{\frac{1}{Q}\sum_{k=1}^{Q}\left(CC_{gt,k} - CC_{res,k}\right)^2}, \quad (7)$$

where Q is the total number of under segmented lines in the text sample, $CC_{gt,k}$ is the number of ground truth lines in the under segmented text line k (equal to one per each line), and $CC_{res,k}$ is the number of resultant connected components in the under segmented text line k. Commonly, under-segmentation represents joined lines error which corresponds to the situation where the sequence of n consecutive lines is considered by the algorithm as an unique line. In that case, and if no other error happens, it is considered that one line in the sequence is correct and the other n - 1 lines of the group are erroneous [11]. $RMSE_{mix}$ is defined as:

$$RMSE_{mix} = \sqrt{\frac{1}{R}\sum_{l=1}^{R}\left(CC_{gt,l} - CC_{res,l}\right)^2}, \quad (8)$$

where R is the total number of mixed lines in the text sample, $CC_{gt,l}$ is the number of ground truth lines in the mixed text line l (equal to one per each line) and $CC_{res,l}$ is the number of resultant connected components in the mixed text line l. This way, $RMSE_{over}$, $RMSE_{under}$ and $RMSE_{mix}$ can be used alongside with or without the measure $RMSE_{seg}$.

Evaluation based on binary classification

Binary classification is based on the theory of the signal detection [18]. Its task is to classify the members of a given set into two groups. The classifying is based on whether they have some property or not. It is represented by a confusion matrix which complies with the text line segmentation as shown in Table 1.

Table 1. Confusion matrix.

Ground Truth \ Resultant	Line	Non-line
Line	*TP* (True Positive)	*FN* (False Negative)
Non-line	*FP* (False Positive)	*TN* (True Negative)

The contents of the above table could be explained in the following way. If some connected components have a ground truth property and the test confirms it as resultant, then those connected components will represent true positives (*TP*). If some connected components do not have a ground truth property, but the test confirms it, then they will represent false positive (*FP*). However, if some connected components have a ground truth property, but the test mistakenly does not confirm it, then they will represent false negative (*FN*). Finally, if some connected components do not have a ground truth property, and the test confirms it, then they will represent true negative (*TN*).

In the context of the classification tasks, all previous statements are used to compare the item classification. Correlation of the previous definitions with the results of the algorithm testing is as follows [1]: *TP* represents segmented text line hits CC_{corr}, *FP* represents the number of the false segmented text line CC_{under} + CC_{mix}, and *FN* represents segmented text line misses CC_{over}. From the above elements the common evaluation measures related to binary classification can be extracted [18-20]. They are: *precision*, *recall* and *f-measure*.

Precision (*P*) is a positive predictive value (*PPV*) defined as the proportion of the true positives against all the positive results (both true positives and false positives). It is a measure of the ability of a system to present only relevant items. In the confusion matrix *TP* represents the number of correct lines, while *TP* + *FP* represents the number of resultant lines given by the algorithm under test. Hence, precision is defined as [18-20]:

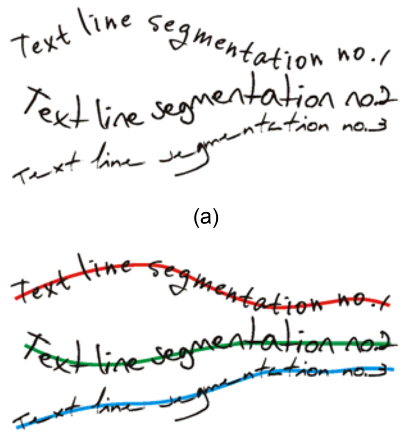
$$P = \frac{TP}{TP + FP} = \frac{\text{Correct Lines}}{\text{Resultant Lines}}. \quad (9)$$

A higher *precision* value means less false positives, and vice versa. *Recall* (*R*) is a measure of the system's ability to present all relevant items. In the confusion matrix *TP* represents the number of correct lines, while *TP* + *FN* represents the number of ground truth lines. Hence, recall is defined as [18-20]:

$$R = \frac{TP}{TP + FN} = \frac{\text{Correct Lines}}{\text{Ground Truth Lines}}. \quad (10)$$

Higher *recall* value means less false negatives, while lower *recall* means more false negatives. *Precision* and *recall* can be combined to produce a single metric known as *f-measure* (*F*), which is the weighted harmonic mean of the *precision* and *recall*. *F-measure* exploits the balance between *precision* and *recall*. It is defined as [18-20]:

$$F = 2\text{x}\frac{P\,\text{x}\,R}{P + R} = \frac{TP}{TP + \frac{FN + FP}{2}}. \quad (11)$$



(a)

(b)

Fig. 4. Text sample: (a) Initial text, and (b) Ground truth segmentation.

Accuracy (*A*) is also used as a statistical measure of how well a binary classification test correctly identifies or excludes a condition. It is the proportion of true results (both true positives and true negatives) in the population. Hence, it is a parameter of the test. *Accuracy* is defined as:

$$A = \frac{TP + TN}{TP + FP + TN + FN}. \quad (12)$$

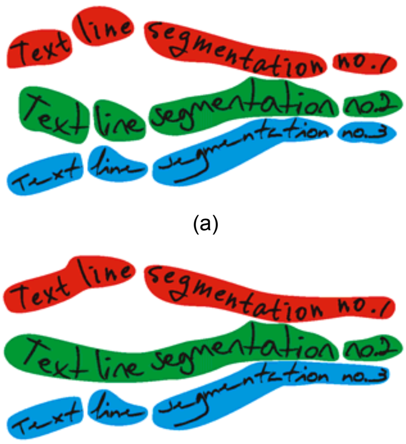
In the extended binary classification method [15], the new measure $RMSE_{seg}$ *given* in eq. (5) is introduced.

Comparative Analysis and Discussion

A comparative analysis was performed by the comparison between two evaluation methods on the real text examples. For a comprehensive comparison many different text examples should be examined. The initial text and its ground truth segmentation are shown in Fig. 4.

Over-segmentation

In the over-segmentation example, the different intensity of these phenomena is shown in Fig. 5.



(a)

(b)

Fig. 5. Over-segmentation: (a) Higher, and (b) Lower intensity.

To analyze and compare different evaluation methods for text line segmentation, theirs behavior as well as their detailed grading and fine distinction in the text segmentation examples will be examined as follows. Comparison of different text segmentation results characterized by over-segmentation is given by both methods in Table 2.

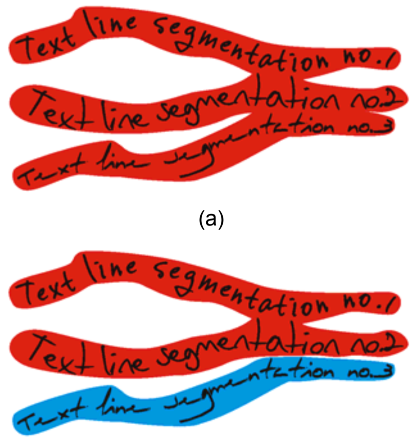
Analyzing the results from Table 2 errors type classification is more common sense. Consequently, error type classification recognized clearly and completely the over-segmentation phenomena by *OSLHR* (1 or 100%) as well as its intensity by $RMSE_{over}$. Binary classification is completely blind. No results have any common sense whatsoever. However, the extension measure $RMSE_{seg}$ [5] clearly distinct the intensity of the over-segmentation. Therefore, there is no evidence that over or mixed segmentation is in place. Furthermore, the higher intensity over-segmentation would give weaker results characterized by higher values of *RMSE*.

Table 2. Evaluation of the over-segmentation for the text sample.

Figure	5.(a)	5.(b)
Text segmentation elements		
CC_{gt}	3	3
CC_{over}	3	3
Line #1	4	2
Line #2	4	2
Line #3	4	3
Error type classification method		
OSLHR	1	1
SLHR, USLHR, MSLHR	0	0
$RMSE_{over}$	4.58	2.31
$RMSE_{under}$, $RMSE_{mix}$	0	0
Binary classification method		
Precision	NA	NA
Recall, F-measure, Accuracy	0	0
$RMSE_{seg}$	4.58	2.31

Under-segmentation

Text sample with different intensity of the under-segmentation phenomena is shown in Fig. 6.



(a)

(b)

Fig. 6. Under-segmentation: (a) Higher, and (b) Lower intensity.

Comparison of different text segmentation results characterized by under-segmentation will be given by both methods in Table 3.

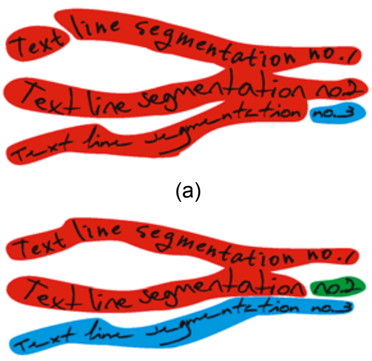
Table 3. Evaluation of the under-segmentation for the text sample.

Figure	6.(a)	6.(b)
Text segmentation elements		
CC_{gt}	3	3
CC_{corr}	-	1
CC_{under}	3	2
Line #1	1	1
Line #2	0	0
Line #3	0	1
Error type classification method		
SLHR, MSLHR	0	0
OSLHR	0	1
USLHR	1	0
$RMSE_{over}$	0	2.31
$RMSE_{under}$	1.42	0
$RMSE_{mix}$	0	0
Binary classification method		
Precision	0	0.33
Recall	NA	1
F-measure	0	0.5
Accuracy	0	0.33
$RMSE_{seg}$	1.42	1.00

Analyzing the results from Table 3 errors type classification is again more common sense. The first four measures *SLHR*, *OSLHR*, *USLHR* and *MSLHR* completely quantify the results. Furthermore, additional three measures $RMSE_{over}$, $RMSE_{under}$ and $RMSE_{mix}$ qualitatively distinct the fine differences between results. However, in the circumstances (a) the binary classification is blinded. The extension measure $RMSE_{seg}$ distinct the intensity measure of the under-segmentation. However, there is no confirmation about the type of error. In the example (b), it seems that binary classification is starting to work. Consequently, qualitative distinction of the error is still missing. At the end, the characterization of any *RMSE* is very valuable for fine tuning of the algorithm.

Mixed-segmentation

Text sample with different intensity of the mixed-segmentation phenomena is shown in Fig. 7.



(a)

(b)

Fig. 7. Mixed-segmentation: (a) Higher, and (b) Lower intensity.

Comparison of different text segmentation results characterized by mixed-segmentation will be given by both methods in Table 4.

Table 4. Evaluation of the mixed-segmentation for the text sample.

Figure	7.(a)	7.(b)
Text segmentation elements		
CC_{gt}	3	3
CC_{corr}	-	1
CC_{under}	3	2
Line #1	2	0
Line #2	0	2
Line #3	2	1
Error type classification method		
SLHR	0	0.33
OSLHR, USLHR	0	0
MSLHR	1	0.67
$RMSE_{over}$, $RMSE_{under}$	0	0
$RMSE_{mix}$	1.53	1.15
Binary classification method		
Precision	0	0.33
Recall	NA	1
F-measure	0	0.5
Accuracy	0	0.33
$RMSE_{seg}$	1.52	1.15

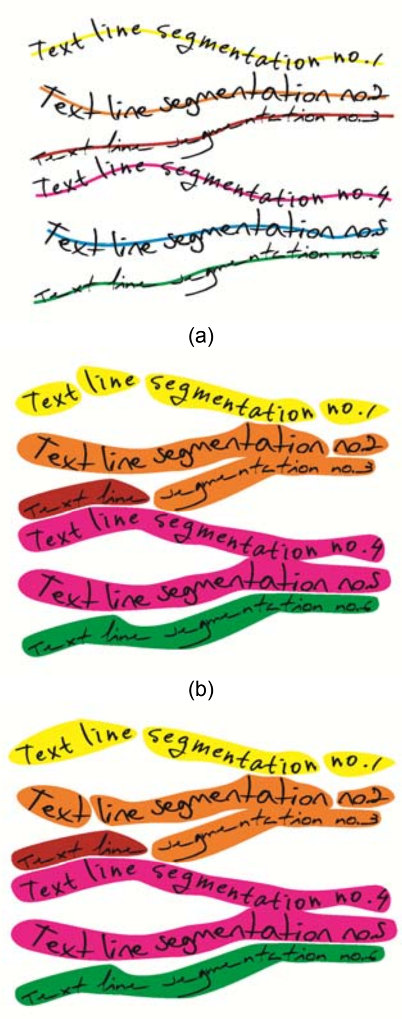
Analyzing the results from Table 4 errors type classification obviously has dominant advantage. Again, *SLHR*, *OSLHR*, *USLHR* and MSLHR completely identify the nature of the algorithm behavior. Additional RMSE measures give the confirmation of the fine differences. Again, in the circumstances (a) the binary classification has no common sense. However, the extended measure $RMSE_{seg}$ is the only measure that quantifies measurement results. In the example (b), it seems that binary classification is beginning to work. Consequently, the fine distinction of the errors is still missing.

Combination of the segmentation

Text samples with combination of different types of segmentation are shown in Fig. 8.

Comparison of the real examples of different types of segmentation results will be given by both methods in Table 5.

Analyzing the results from Table 5 errors type classification gives more measure values which pin-point the real circumstance. *SLHR*, *OSLHR*, *USLHR* and *MSLHR* globally identify the algorithm behavior. Additional *RMSE* measures give the local observation. Hence, the each *RMSE* pinpoint the real difference between segmentation results. Again, the binary classification method is blind to the different segmentation results. It gives only global identification of the segmentation, without any local information. In addition, $RMSE_{seg}$ makes up the evaluation. However, it is globally oriented measure which can not distinct the local differences. Hence, the fine distinction of the errors is lacking.



(a)

(b)

(c)

Fig. 8. Mixed-segmentation: Combined segmentation: (a) Ground truth segmentation, (b) Segmentation 1, (c) Segmentation 2.

Table 5. Evaluation of the combined segmentation for the text sample.

Figure	8.(a)	8.(b)
Text segmentation elements		
CC_{gt}	6	6
CC_{corr}	1	1
CC_{over}	1	1
CC_{under}	2	2
CC_{mix}	2	2
Line #1	4	3
Line #2	2	3
Line #3	2	2
Line #4	1	1
Line #5	0	0
Line #6	1	1
Error type classification method		
SLHR, OSLHR	0.17	0.17
USLHR, MSLHR	0.33	0.33
$RMSE_{over}$	3	2
$RMSE_{under}$	0.71	0.71
$RMSE_{mix}$	1.22	1.73
Binary classification method		
Precision	0.2	0.2
Recall	0.5	0.5
F-measure	0.28	0.28
Accuracy	0.17	0.17
$RMSE_{seg}$	2.12	2.58

Discussion

The evaluation based on error type contains seven distinct measures: *SLHR*, *OSLHR*, *USLHR*, *MLHR*, $RMSE_{over}$, $RMSE_{under}$ and $RMSE_{mix}$. Their interpretation is clear and unmistakable. The last three measures are very useful for the fine tuning of the segmentation results that have to be reached. Obviously, the evaluation based on error type is more clear and remarkable. In contrast, the evaluation based on the binary classification has only four distinct measures: *precision*, *recall*, *f-measure* and *accuracy*. In addition, the third one is the harmonic mean of the first two. Nevertheless, this evaluation process includes more statistical measures. In [15], the evaluation based on binary classification is improved by additional measure $RMSE_{seg}$. This measure contributes to the fine evaluation between results. However, it has no capability to separate and characterize different types of errors which is of great importance for the algorithm optimization. $RMSE_{over}$, $RMSE_{under}$ and $RMSE_{mix}$ can be used alongside with or without the measure $RMSE_{seg}$. Still, the method based on error type classification has numerous advantages.

Conclusions

The paper presents two extended methods for the evaluation of the algorithm for the text line segmentation. The first method is based on the text line segmentation error terms. In its extended version, it incorporates seven distinct measures. These measures are divided into globally and locally oriented. However, they are strongly linked and mutually amended. The second method, which is well known and most often used, is based on the binary classification linked with signal detection theory. It consists of four distinct measures. Its extended version incorporates additional measure $RMSE_{seg}$. Both methods have been explored in the real text line segmentation circumstances. However, due to the seven measures that characterize the evaluation process, the error type method has clear advantages. It is especially true for small differences in segmentation process which is correctly pin-point by this method. Hence, it is proved as correct, solid and robust evaluation method.

REFERENCES

[1] Du, X., Pan, W., Bui, T. D., Text line segmentation in handwritten documents using Mumford–Shah model, *Pattern Recognition*, 42 (2009), No. 12, 3136–3145.

[2] Likforman Sulem, L., Zahour, A., Taconet, B., Text line segmentation of historical documents: a survey, *International Journal on Document Analysis and Recognition*, 9 (2007), No. 2, 123–138.

[3] Amin, A. and Wu, S., Robust skew detection in mixed text/graphics documents, Proceedings of International Conference on Document Analysis and Recognition - ICDAR'05, (2005), Seoul, Korea, 247–251.

[4] Bukhari, S. S., Shafait, F., Breuel, T. M., Script-Independent handwritten textlines segmentation using active contours, Proceedings of International Conference on Document Analysis and Recognition - ICDAR'09, (2009), Barcelona, Spain, 446–450.

[5] Yi, L., Yefeng, Z., Doermann, D., Jaeger, S., Script-Independent text line segmentation in freestyle handwritten documents, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (2008), No. 8, 1313–1329.

[6] Marti, U. V. and Bunke, H., The IAM-database: an English sentence database for off-line handwriting recognition, *International Journal on Document Analysis and Recognition*, 5 (2002), No. 1, 39–46.

[7] Basu, S., Chaudhuri, C., Kundu, M. et al., Text line extraction from multi-skewed handwritten document, *Pattern Recognition*, 40 (2007), No. 6, 1825–1839.

[8] Sarkar, R., et al., CMATERdb1: a database of unconstrained handwritten Bangla and Bangla–English mixed script document image, *International Journal on Document Analysis and Recognition*, 14 (2011), No. 1, 25–33.

[9] Gatos, B., Stamatopoulos, N., Louloudis, G., ICDAR2009 handwriting segmentation contest, *International Journal on Document Analysis and Recognition*, 14 (2011), No. 1, 1–13.

[10] Louloudis, G., Gatos, B., Pratikakis, I., Halatsis, C., Text line and word segmentation of handwritten documents, *Pattern Recognition*, 42 (2009), No. 12, 3169–3183.

[11] Sanchez, A., et al., Text line segmentation in images of handwritten historical documents, Proceedings of the First Workshops on Image Processing Theory Tools and Applications - IPTA 2008, (2008), Sousse, Tunisia, 1–6.

[12] Brodić, D., Milivojević, D. R., Milivojević, Z., An Approach to a comprehensive test framework for analysis and evaluation of text line segmentation algorithms, *Sensors*, 11 (2011), No. 9, 8782–8812.

[13] Brodić, D., Milivojević, D. R., Milivojević, Z., Basic test framework for the evaluation of text line segmentation and text parameter extraction, *Sensors*, 10 (2010), No. 5, 5263–5279.

[14] Brodić, D., Basic experiments set for the evaluation of the text line segmentation, *Przegląd Elektrotechniczny*, 86 (2010), No. 11b, 353–357.

[15] Brodić, D., Methodology for the evaluation of the algorithms for text line segmentation based on extended binary classification, *Measurement Science Review*, 11 (2011), No. 3, 71–78.

[16] Thulke, M., Märgner, V., Denge, A., A General approach to quality evaluation of document segmentation results. In Le, S. W. and Nakano, Y. (eds.) Document analysis systems: theory and practice. LNCS, 1655, Berlin-Heidelberg: Springer-Verlag, (1999), 43–57.

[17] Gonzales, R. C., Woods, R. E., Digital Signal Processing, 2nd ed., (2002), Prentice-Hall.

[18] Swets, J. A., Measuring the accuracy of diagnostic systems, *Science*, 240 (1988), No. 4857, 1285–1293.

[19] Qian, X., Liu, G., Wang, H., Su, R., Text detection localization and tracking in compressed video, *Signal Processing: Image Communication*, 22 (2007), No. 9, 752–768.

[20] Bukhari, S. S., Shafait, F., Bruesl, T. M., Adaptive binarization of unconstrained hand-held camera-captured document images, *Journal of Universal Computer Sc*ience, 15 (2009), No. 18, 3343–3363.

***Authors**:*
Ass. Prof. dr Darko Brodić, University of Belgrade,Technical Faculty in Bor, V.J. 12, 19210 Bor,Serbia, E-mail: dbrodic@tf.bor.ac.rs
Prof. dr Zoran N. Milivojević, Techical College Niš, A. Medvedeva 20, 18000 Niš, Serbia, E-mail: zoran.milivojevic@jotel.co.rs
Dr Dragan R. Milivojević, Institute for Mining and Metallurgy, Zeleni Bulevar 5, 19210 Bor, Serbia, E-mail: dragan.milivojevic@irmbor.rs