7

Core Region Extraction for Off-line Unconstrained Handwritten Words

Zone extraction is acclaimed as a significant preprocessing step in handwriting analysis. In this chapter, a new method for separating ascenders and descenders from an unconstrained handwritten word is proposed. The extracted core-region is quite accurate, as demonstrated by the experimental comparisons with the *state-of-the-art* core region extraction methods. In the following chapter, we make use of the extracted core-region to extract features for writer identification. This chapter is organized as follows. Section 7.1 presents the challenges and need of core-region detection in document image analysis. In Section 7.2 we discuss the related work. Section 7.3 focuses on the proposed methodology and provides a detailed analysis of the steps involved. Section 7.4 presents the experimental results indicating the performance of the proposed method compared to other state-of the-art methods. Section 7.5 presents the concluding remarks.

7.1 INTRODUCTION

An individual's handwriting is influenced by several factors. The handwriting may get deteriorated due to writing in surge, shuddering of hands due to age, presence of different fonts, letter casing, and inconsistent position on the page. This results in an assortment of handwriting challenges like those shown in Figure 7.1. Alongside, there are other challenges like words written

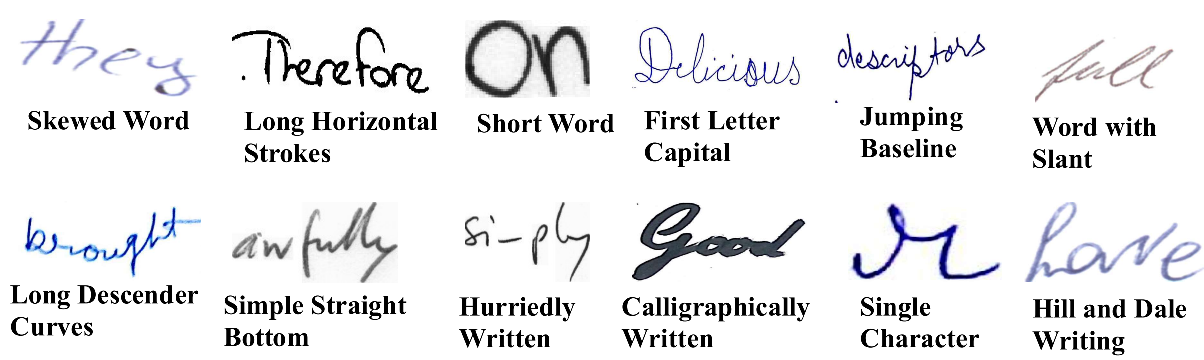


Figure 7.1 : Variety of handwriting challenges for core-region extraction.

in all upper case, cursive handwriting, calligraphic words, title case words, short words such as \`on', \`in' etc. These complex variations in handwriting effectively lower the performance of any handwriting analysis system. Text normalization maneuvers such difficulties by transforming the varied handwritten text into a single canonical form that leads to optimal feature extraction for similar classes. Zoning is one such technique which normalizes the text and makes it suitable for optimized feature extraction. In zoning, a word is mainly divided into three regions by means of lines called as upper and lower reference lines or baselines. The region above the upper reference line is termed as the ascender zone; and the region below the lower reference line is termed as the descender zone. The area enclosed between the upper and lower reference/baselines is termed as core-region. Figure. 7.2 details the respective zones with lines in a word.

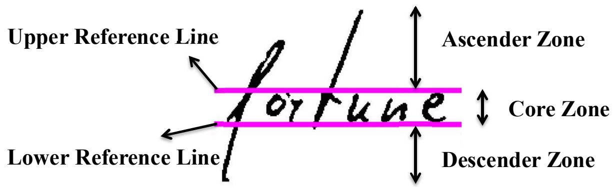


Figure 7.2 : A word divided into its corresponding zones.

As noted by [Bozinovic and Srihari, 1989; Castro-bleda *et al.*, 2011; Stahlberg and Vogel, 2015], erroneous determination of the core zone can lead to loss of information about the shape of the character, which affects the efficacy of handwriting analysis systems. Moreover, in a word the height of the core zone corresponds to the height of the lower case characters of the word, which is used as a feature normalization parameter [Pastor-Pellicer *et al.*, 2014]. Information about the ascender, descender and core zone has been widely applied in the areas of handwriting recognition [Bozinovic and Srihari, 1989; Kavallieratou *et al.*, 2002], script and style identification [Slimane *et al.*, 2014], writer identification [Schlapbach and Bunke, 2005; Gazzah and Ben Amara, 2007; Slimane and Margner, 2014], printed and handwritten text segmentation [Kavallieratou *et al.*, 2004; Echi and Saidani, 2014], skew and slant correction [van Beusekom *et al.*, 2009], ink beautification [Miyao and Maruyama, 2012], and multi-font character recognition [Jung *et al.*, 1999].

Separating the zones for uneven handwriting is a difficult task. Traditionally, methods like horizontal projection profile and its variants were used to mark the zones of a word. These methods were strongly affected by the presence of long horizontal strokes and hill-and-dale writing, and consequently were limited to application on constrained text. Other methods like polygon approximations and line regression were also used to find the reference lines of the word but these were specific to Arabic text [Maddouri *et al.*, 2008; Al-Shatnawi and Omar, 2008]. Collectively, the studies outline a requirement of an approach that can handle variegated handwriting while extracting zones in a word.

This paper seeks to remedy these problems by presenting a new method for separating ascenders and descenders regions from an unconstrained handwritten word and identify its core-region. The method is able to handle many of the challenges shown in Figure 7.1. The method extracts two word envelops and selects points on the envelops that are likely to belong to the core region. The proposed algorithm has been tested on CVL, ICDAR-2013, ICFHR-2012, and IAM benchmark datasets of handwritten words written by multiple writers for English.

We also created our own dataset penned by 2 writers comprising 100 words that includes words with jumping, curved, skewed and hill-dale baselines. It also comprises calligraphic words, short and long words and slanted words. Because of the unavailability of the ground truth for core-region, we generated it ourselves for all the datasets. To measure the correctness of the estimated core-region, the results are compared with the ground truth as suggested in [Pechwitz *et al.*, 2012; Stahlberg and Vogel, 2015]. Our work reports an accuracy of 90.16% for correctly identifying all the three zones on a total of 17,100 Latin words written by 802 individuals from all the datasets.

7.2 RELATED WORK

Traditionally, horizontal projection profile (HPP) has been applied to find the reference lines. HPP shows the counts of the foreground pixels for each row (y coordinate) on the y-axis.

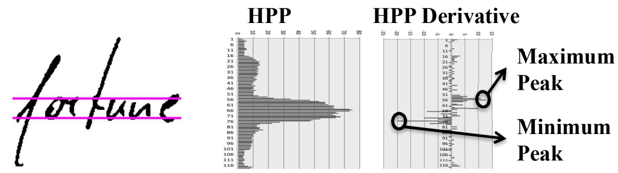


Figure 7.3 : Zoning by using Horizontal Projection Profile Method

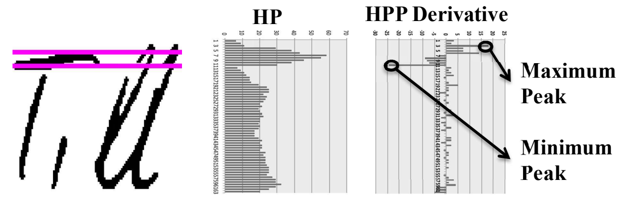


Figure 7.4 : Erroneous detection of reference lines due to longer horizontal strokes by Horizontal Projection Profile method.

The maximum and minimum peaks of the first derivative of the HPP are selected as locations for the upper and lower reference lines respectively [Bozinovic and Srihari, 1989]. Figure 7.3 shows the reference lines for the word *fortune* (from CVL dataset). Unfortunately, HPP is not effective for core zone extraction due to huge diversity in handwriting. Figure 7.4 illustrates an example where because of the long stroke of the capital letter *T* incorrect core-region for the word *Till* (from CVL dataset) is extracted. With a need to provide better results for a wide variety in handwritings, a considerable amount of literature has been published on core-region extraction for Latin script. Usually, the large horizontal strokes cause erroneous maximum and minimum peaks in the HPP. [Bozinovic and Srihari, 1989] proposed a modification to the HPP method that correctly estimates the core-region for the words with long horizontal strokes. They ignore *k* maximum peak values in the HPP with an intuition that those *k* peaks belong to long horizontal strokes. A new maximum peak (*max*) is selected from the remaining values of the HPP. Using this *max* and by means of a threshold the shoulders of the HPP are determined that mark the upper and lower reference lines (Figure 7.5). The method works fairly well for the words whose reference lines are well aligned with the horizontal.

[Côté *et al.*, 1998] compute several histograms of foreground pixel counts in different directions and find the entropy associated with each of them. The histogram having the lowest entropy is considered to correspond to the writing direction. This histogram is selected and further

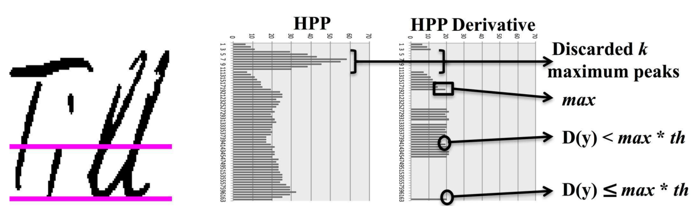


Figure 7.5 : The peaks of the horizontal stroke are discarded in the HPP derivative while the peaks of the core-region are preserved [Bozinovic and Srihari, 1989].

processed to apply heuristics for separating the core-region. The method worked well for skewed and slanted words. However, words with calligraphic handwriting and upper case words still pose a challenge.

[Vinciarelli and Luettin, 2001] also computed distribution of the foreground pixel counts for the image rows. They noted that this distribution is bimodal where one mode corresponds to the long strokes while the other to the core region. They point out that the bimodal histogram is quite robust to contributions arising due to long horizontal strokes. A threshold selected using the Otsu's algorithm is applied on the HPP and the region exceeding this threshold (in terms of foreground pixel count) is marked as core-region. To find the lower reference line the lowest point on each column of the image is selected and the least square method is used to fit those points. The method yielded satisfactory zone segmentation for the purpose of normalizing slant and skewed handwriting. Nonetheless, for hill and dale handwriting and for words with long descender curves, the method does not produce substantial results.

[Blumenstein *et al.*, 2002] identified the upper baseline as the highest row (in the upper portion of the word) with foreground pixel count exceeding the average count. To get the lower baseline two candidate rows are identified (in the lower portion), one containing the minimum number of foreground pixels (minPixY) and the other one as the lowest row containing more than the average number of foreground pixels (avgPixY). If minPixY is too close to avgPixY then minPixY is chosen as the lower baseline, otherwise avgPixY is chosen. They designed their method for slant and skew correction of words; hence it is well suited for such handwritings. However, considerable results were not produced for intricacies like hill and dale handwriting and words with long descender curves.

[Rehman *et al.*, 2009] devised a threshold selection method in which the HPP values are sorted and a fixed point at three fourth index is selected. They call it as *quantile*, which acts as a threshold to find the reference lines. The highest peak is selected from HPP and is marked as middle line. On the HPP, examining the y-projection starting from the first row toward the middle line, the y-projection with the number of pixels greater or equal to the threshold (*quantile*) is detected as upper-baseline. Similarly, starting from the last row toward the middle line, the y-projection greater than or equal to *quantile* is marked as lower reference line. The method correctly finds the ascenders and descenders for cursive handwriting without having slant and skew.

[Papandreou and Gatos, 2014] introduced a reinforced horizontal black run profile histogram which focused on the horizontal black runs of the word image. They demonstrated that the core-region constitutes more black to white transitions and hence will have multiple horizontal black runs. The non-core region has smaller runs of foreground pixels because vertical strokes in non-core region have no significant width and the horizontal strokes in the non-core region are sparse. A Boolean HPP array is constructed from horizontal run-length profile histogram that will have either 1 or 0, based on comparison with a calculated threshold. The upper and lower reference lines are marked according to the maximum number of successive 1s in the Boolean array. Their method is well suited for slant and skewed handwriting.

Deviating from the histogram-based methods [Pastor-Pellicer *et al.*, 2014] used an HMM hybridized with ANNs to identify the reference lines. The classifier is applied column-wise and each pixel in the column is classified into one of the three classes indicating belongingness to ascender, descender or core zone. The method can find the zones only for a slant and skew corrected text-line.

Apart from Latin text there is Arabic script which has a strong baseline which is exploited for feature extraction by handwriting recognition methods [Stahlberg and Vogel, 2015]. Methods like HPP, word skeleton based polygon approximations, contour representations and PCA have been used to find the baseline in Arabic handwriting. A comparative overview of different methods for Arabic lower and upper baseline detection can be found in [Al-Shatnawi and Omar, 2008] and [Al-Shatnawi and Omar, 2009].

Apart from the methods described in [Al-Shatnawi and Omar, 2008, 2009], [Boubaker *et al.*, 2009] presented a method to extract baselines for off-line short handwriting with an assumption that the text is already slanted at a specific angle θ. The algorithm detected straight and curved lower baseline for short Arabic handwritten words.

[Stahlberg and Vogel, 2015] estimated lower baseline for Arabic handwritten words. They cut many vertical strips of specified width on the word image and find the most dense (best strip) foreground pixels strip. The bottom of that selected strip is chosen to be the lower baseline for the given word. They also worked on the images with fluctuating lower baseline. They divided the image into sub parts and then applied their method individually and merged the results. [Boukharouba, 2016] tracked the lower baseline in text lines by deskewing the images using a randomized Hough transform and then by applying HPP method to search for the lower baseline.

It is clear that the previous work on Latin text has major focused on skewed and slanted words comprising long horizontal and vertical strokes. However, variations such as those shown in Figure 7.1 need to be addressed for core-region extraction. The method presented in this paper aims to address these challenges to achieve robust core-region detection for Latin words.

7.3 METHODOLOGY

We extract a set of word envelops for the full word image and its cropped version. More specifically, we construct two upper envelops, one (denoted U_f) for the full word image and the other (U_c) for a cropped word image which retains the lower 2/3 portion and crops out the upper 1/3 portion. Likewise we construct two lower envelops, one (L_f) for the full image and the other (L_c) for a cropped image which retains the upper 2/3 portion and crops out the lower 1/3 portion. The steps are as follows.

1. Identify sample points on the 4 envelops U_f, U_c, L_f, L_c.
   To sample the upper envelops U_f and U_c the consecutive pixels on the upper profile are processed in sets comprising 3 consecutive pixels. From each set we retain the pixel with the largest y coordinate (see Figure 7.6,7.7).

   To sample the lower envelops L_f and L_c we follow an identical process except that we retain the pixels with the smallest y coordinate.

2. Identify extrema points on the sampled envelops. For the sampled upper envelops U_f and U_c we identify the maxima points, denoted as sets MU_f and MU_c respectively, using the following steps:

   a) Consider each sample pixel p_n in the context of its two neighbors p_{n-1} and p_{n+1} and examine the placement of p_{n-1}, p_n, p_{n+1} to see if it matches with one of the topological patterns A, B or C as shown in Figure 7.8.

   b) If there is a match then the pixel p_n is declared as a local maximum point (see Figure 7.9).

   c) The first and the last point in the sequence are also included in the sets MU_f and MU_c.

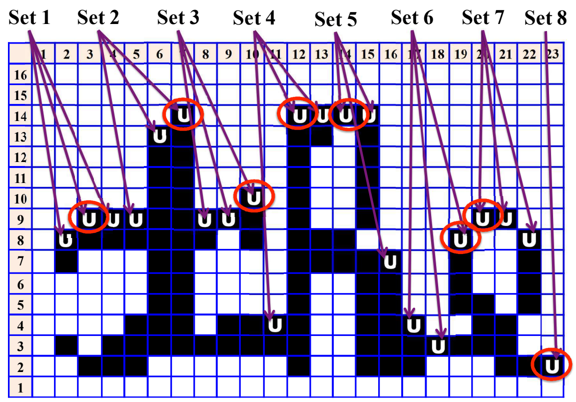


Figure 7.6 : Contiguous set of pixels are drawn on the upper envelop. The pixel marked as **U** belongs to the upper envelop.

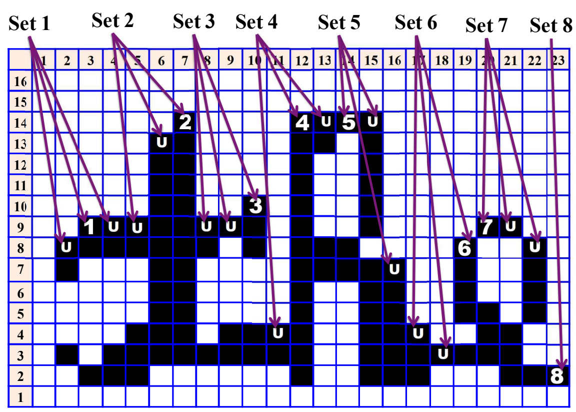
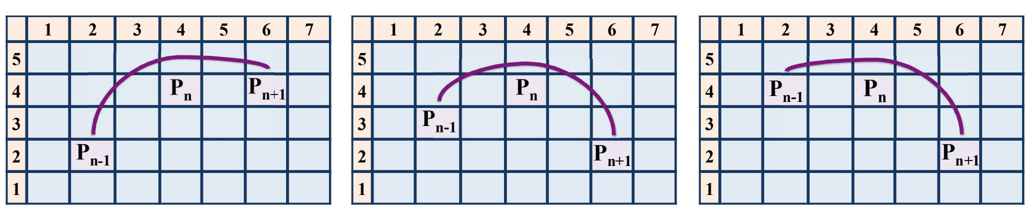


Figure 7.7 : The pixels numbered from **1** --- **8** represents the sampled sequence of pixels on the upper envelop.



(a) Left Half-Bell Shaped Pixel Curve (b) Full-Bell Shaped Pixel Curve (c) Right Half-Bell Shaped Pixel Curve

Figure 7.8 : Topological Patterns For Upper Reference Lines.

For the sampled lower envelops L_f and L_c we identify the minima points, denoted as sets ML_f and ML_c respectively, using identical steps with modified topological patterns shown in Figure 7.10.

3. The extrema pixels may lie on the ascender/descender or the core zone. Therefore the sets

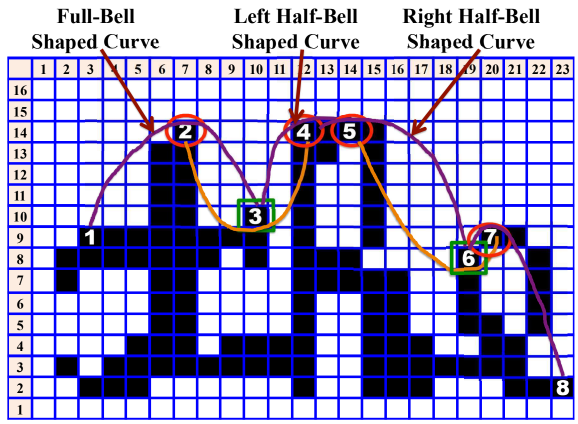
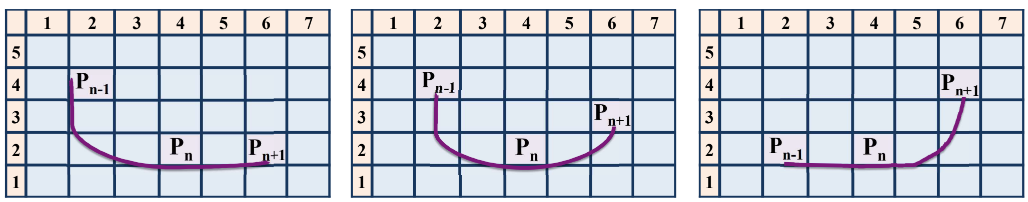


Figure 7.9 : The pixels **2, 4, 5** and **7** are the maximal points selected when matched with the Topological Patterns for the upper reference line.



(a) Left Half-Inverted Bell Shaped Pixel Curve (b) Full-Inverted Bell Shaped Pixel Curve (c) Right Half-Inverted Bell Shaped Pixel Curve

Figure 7.10 : Topological Patterns For Lower Reference Lines.

of maxima pixels and minima pixels are then processed to remove the outlier points, i.e. the points possibly lying on the ascender or descender and retain the inliers, i.e. the points possibly lying on the core region. For each of the sets MU_f and MU_c the following steps are taken for inlier selection.

a) Calculate the mean and the standard deviation of the y coordinates of the maxima points.

b) For each maxima point:

   i. Calculate the deviation of its height (y-coordinate) with respect to the mean height (see Figure 7.11).

   ii. If the height deviation is less than the standard deviation, the pixel is considered as an inlier.

   The sets of inliers obtained for MU_f and MU_c are denoted as CU_f and CU_c respectively.

A similar inlier selection is done for the minima pixels in each of the sets ML_f and ML_c to give the sets CL_f and CL_c.

It is to be noted that the sets CU_f, CU_c, CL_f and CL_c may include some pixels which are not

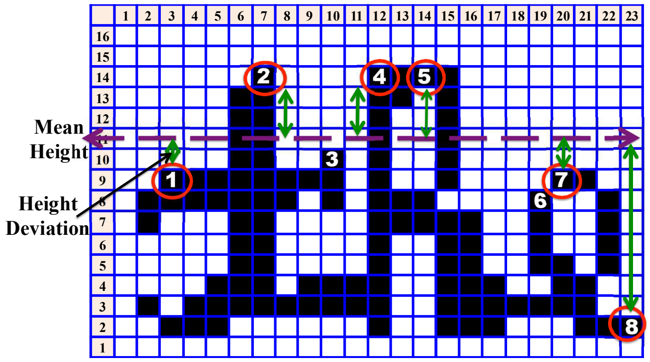


Figure 7.11 : Partitioning the computed maxima points into two sets of core and non-core region maximal points. The first and the last maximal points i.e. **1** and **8** are also included.

on the core zone. Therefore, a final selection is made for the sets which are likely to contain a larger number of points on the core zone.

Points on the upper envelop of the core zone is obtained from one of the sets CU_f and CU_c using the following steps:

a) For each of the two identified inlier sets CU_f and CU_c, compute the difference between the maximum and minimum y-coordinates of points in that set.

b) The set CU_f or CU_c which computes a smaller difference is selected as the upper envelop of the core zone.

Analogous steps are followed to select one of the sets CL_f and CL_c as the lower envelop of the core zone. The region inside the upper and lower envelop will give the core-region of the respective image. It must be noted that the obtained upper and lower baselines are not straight lines. Instead we obtain a sequence of points which when connected form a kind of envelop for the core region.

7.4 EXPERIMENTAL DETAILS AND RESULTS

In order to test the proposed methodology, we investigated the following artifacts of handwriting: long horizontal strokes, skewed words,short words, first letter capital, hill and dale writing, jumping baselines and words with long descender curves. We manually created ground truth for four different benchmark datasets to measure the effectiveness of our method. We also created our own dataset to experiment on larger complexities of handwriting. Table 7.1 provides the summary of the composition of Dataset.

Table 7.1 : Composition of Dataset along with the achieved accuracies.

Data Set	**ICDAR 2013**	**ICFHR 2012**	**CVL**	**IAM**	**Our Dataset**	**Combined**
Words	5000	2000	6200	3800	100	17100
Writers	250	100	310	140	2	802
Accuracy	**95.89%**	**94.09%**	**92.35%**	**96.49%**	**72.02%**	**90.16%**

The process of ground truth preparation involves marking the core-region envelop using an interactive interface in which the user clicks a set of points on the upper and lower envelop..

The set of line segments connecting the consecutive marked points is taken as the ground truth envelop. To assess the variability in the ground truth markings by different users, we asked 10 users to mark the ground truth core regions for 5 word images. It was observed that there is a variation of up to 20% in the location of pixels (on core-region envelop) in each column of ground truth. Therefore, we conservatively assume that a mismatch with the ground truth core region up to 20% can be considered as acceptable.

As suggested by [Pechwitz *et al.*, 2012], the correctness of the extracted core-region depends on the number of foreground pixels matching with the ground truth core-region. We analyze the performance of our method by computing the maximum percentage of foreground pixels matching with the ground truth core region for all the word images in the datasets. We plot a histogram indicating the proportion of word images whose foreground pixels mismatch percentage lies in intervals [0-5%], [5-10%], [10-20%], [20-30%], [30-40%], [40-50%], [60-70%], [70-80%] (see Figure 7.12). Considering up to 20% foreground pixels mismatch to be acceptably correct, i.e. an

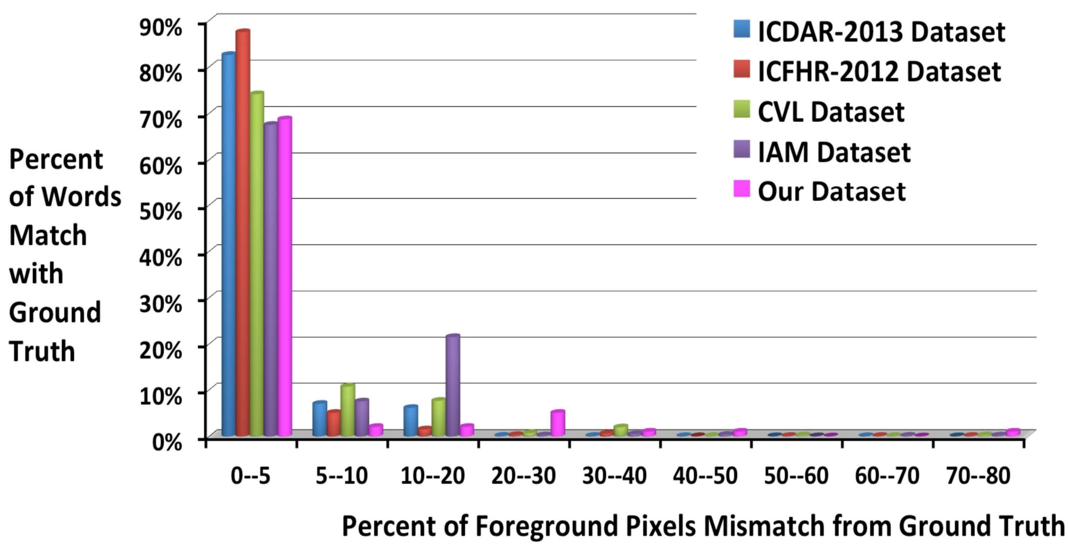


Figure 7.12 : Histogram indicating the percentage of word images having core-region area mismatch with its corresponding ground truth core-region for specified percentage of pixels mismatch.

indicator of correct core region extraction we also compare the accuracy that is obtained with other methods.

Table 7.2 shows the comparison of results obtained by considering correct core region detection to be one having within 20% foreground pixels mismatch.

The results obtained from our method achieve the highest accuracy for all the datasets. However, it is worth mentioning that the reference lines obtained by other techniques are straight lines and our ground truth specifies the core region as bounded by envelops. Therefore, our method yields a higher accuracy. Figure 7.13 shows the results achieved by our method in comparison with ground truth.

A major cause of failure cases (incorrect detection) for our method is the presence of skew exceeding 30 degrees. This happens because a fixed proportion of the word image gets cropped out. This can result in cropping of essential parts of the word image, thereby giving envelops with some points sampled from ascenders or descenders.

Table 7.2 : Performance accuracies of our core-region detection method for handwritten words in comparison with the *state-of-the-art* methods on different datasets.

	ICDAR 2013	ICFHR 2012	CVL	IAM	Our Dataset	Combined
Bozinovic and Srihari [1989]	75.56%	74.90%	72.61%	76.45%	56.59%	71.22%
Vinciarelli and Luettin [2001]	50.50%	55.80%	58.75%	52.03%	44.09%	52.23%
Blumenstein *et al.* [2002]	56.89%	49.51%	59.21%	50.53%	45.23%	52.26%
Rehman *et al.* [2009]	69.50%	68.67%	67.67%	72.09%	49.89%	65.56%
Papandreou and Gatos [2014]	90.02%	88.99%	88.50%	93.09	66.08%	85.14%
Our Method	**95.89**%	**94.09**%	**92.35%**	**96.49**%	**72.02**%	**90.16**%

7.5 CONCLUSION

In this chapter we developed a new method for extracting the core-region from a handwritten word for Latin text. Our method provides significant results for handwriting challenges like presence of long horizontal strokes e.g. \`t',\`F',\`T' etc., hill-dale baseline, jumping baseline, words with long descender curves, cursive handwriting, calligraphic words, title case words, etc. We also have worked on other challenges like words written in all upper case, very short words as \`on', \`in' etc. Our work provides correct results even for moderately skewed image. Future work would consider finding core regions for word images with higher skew by developing automatic threshold selection techniques.

In the next Chapter 8 we address the task of writer identification for handwritten words.

…

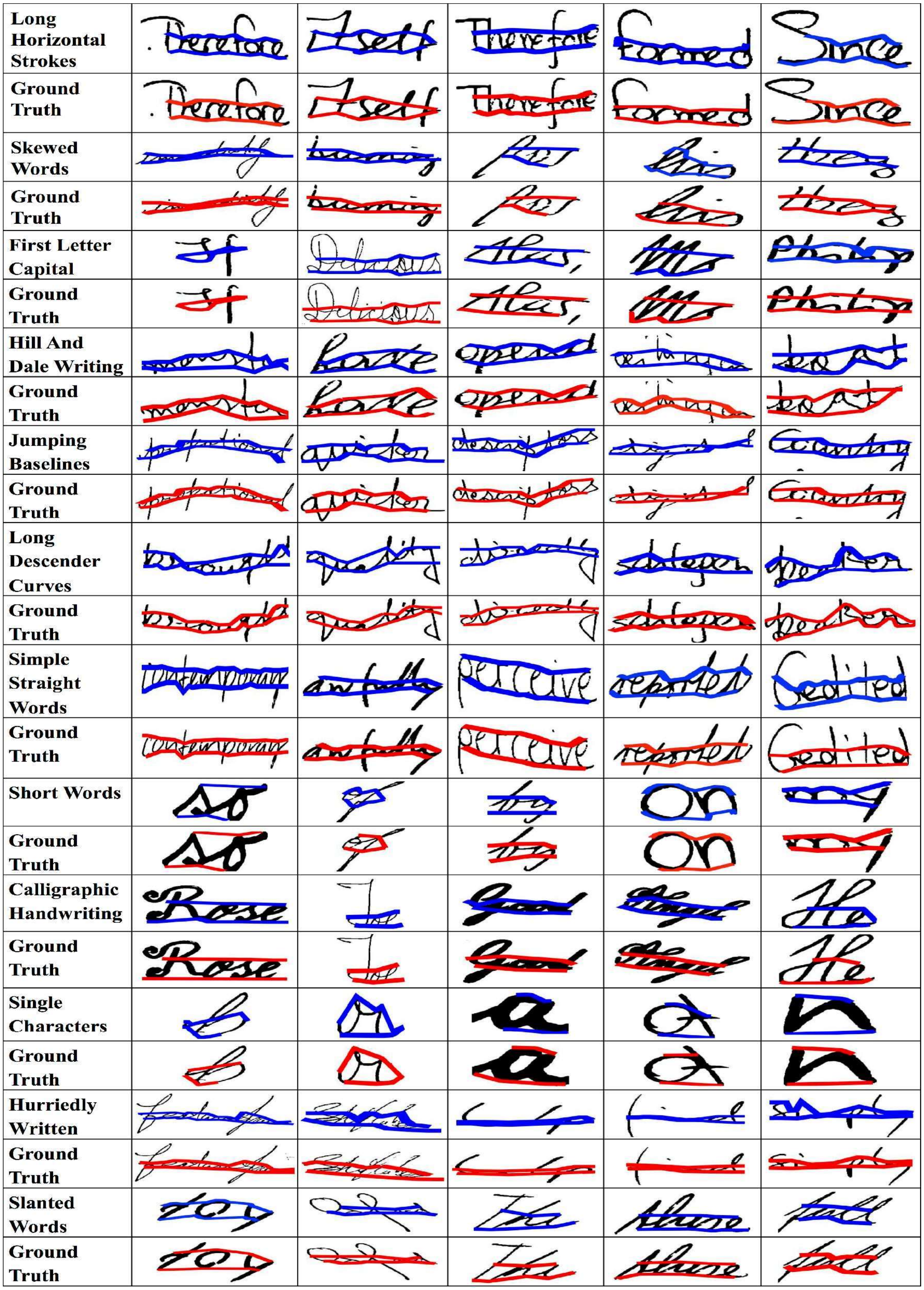


Figure 7.13 : A few sample results of our proposed method to identify core regions in handwritten words. The region between blue lines indicates the core region obtained by our proposed method while the region between red lines indicates the core region marked as ground truth.