A Tracking Approach for Text Line Segmentation in Handwritten Documents

Insaf Setitra, Zineb Hadjadj and Abdelkrim Meziane
Research Center on Scientific and Technical Information, CERIST, Algiers, Algeria
{*isetitra, zhadjadj, ameziane*}*@cerist.dz*

Keywords: Tracking, Connected Components, Angle Minimization, Segmentation, Line, Handwritten, Prediction.

Abstract: Tracking of objects in videos consists of giving a label to the same object moving in different frames. This labelling is performed by predicting position of the object given its set of features observed in previous frames. In this work, we apply the same rationale by considering each connected component in the manuscript as a moving object and to track it so that to minimize the distance and angle of of the connected component to its nearest neighbour. The approach was applied to images of ICDAR 2013 handwritten segmentation contest and proved to be robust against text orientation, size and writing script.

1 INTRODUCTION

Text line segmentation is considered a non-trivial task to solve the field of handwritten document recognition (Stamatopoulos et al., 2013). By and large, many challenges can occur when segmenting handwritten document images such as skew in text lines and adjacency of text lines. To solve some of the text line segmentation challenges, literature count a variety of text line segmentation techniques. Li et al. (Li et al., 2008) proposed an approach for handwritten text line segmentation using level sets. Goto and Aso (Goto and Aso, 1999) proposed a local linearity based method to detect text lines in English and Chinese documents. In the method proposed by Hones and Litcher (Hönes and Lichter, 1994), text lines are generated by expanding the line anchors of the document image. The previously cited methods cannot handle variable sized text, which is the main drawback.

Roy et al. (Roy et al., 2012) proposed text line extraction using foreground and background informations. Louloudis et al. (Louloudis et al., 2007) used a block-Based Hough Transform for text line extraction. In the method proposed by Loo and Tan (Loo and Tan, 2002) the irregular pyramids are used for text line segmentation. Recently, Bukhari et al. (Bukhari et al., 2008) proposed a line segmentation approach for camera-based warped documents using active contour models. Gatos et al. (Gatos et al., 2007) proposed an algorithm based on text line and word detection for warped documents. Bai et al. (Bai, 2008) used a traditional perceptual grouping-based algorithm for extracting curved lines. Pal and Roy (Pal and Roy, 2004) proposed a head-line based technique for multi-oriented (printed in several orientations) and curved text lines extraction from Indian documents. In other work, Pal et al. (Pal et al., 2003) developed a system for English multi-oriented text line extraction estimating the equation of the text line from the character information.

Although cited approaches ware competitive, they still lack universality and the problem of text lines especially in curved document remains open. This paper describes a new approach inspired of tracking works to detect lines in handwritten document images.

Basically, tracking is the process of following objects through an image sequence (Mitiche and Aggarwal, 2014). The earliest methods were focussed on following the trajectory of a few feature points through the sequence. Examples include Kalman filter (Bar-Shalom, 1987) (Broida and Chellappa, 1986) (Boykov and Huttenlocher, 2000) and are applied in areas such as in (Li et al., 2010) and (Zheng et al., 2012).

In our approach, each cluster of connected pixels (which can be a word or a part of a word) is considered as a moving object and it seeks for its best match which satisfies a trajectory angle. Position of the cluster (defined as a connected component in the remaining of the paper) is predicted using prevision position and angle between previous and current observation must be minimized in order to have a best match of

Setitra, I., Hadjadj, Z. and Meziane, A.
A Tracking Approach for Text Line Segmentation in Handwritten Documents.
DOI: 10.5220/0006199001930198
In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)*, pages 193-198
ISBN: 978-989-758-222-6

the current cluster. Clusters are tracked and matched pair to pair until the end of the document and lines are result of that matching. We benefit from tracking rationale where we predict position of clusters according to their history, and avoid tracking issues since we do not count on any feature of the cluster which is obvious since clusters (words or parts of words in the handwritten document image) need not necessarily to be similar in shape or be a weak deformation of their best matches.

We explain more deeply our approach in section 2. We give a preliminary set of tests and criticize them in section 3 and conclude in section 4.

2 OUR APPROACH

Input of our algorithm are binary manuscript images, the first step is then not to segment images, but instead, extract directly connected components that will be tracked over the manuscript image. Connected components are in our case sets of pixels that have a 4 connectivity link. We also remove small regions of connected components since they can false our line segmentation. Examples of such regions include dots (dark pink dots in figure 1.). As a result of this first step, we have a set of connected components each represented by row and column indices of the connected component center (yellow dots in figure 1.). In order to group each set of connected components into a line, we first perform a binary matching of each pair of connected components so that each pair is in the same line. Connected components are assumed to be in the same line if they are spatially close to each other and if the angle they produce with their origin is below a certain threshold.

More specifically, Let $X = \{x^1, x^2, \ldots x^n\}$ be a set of connected component centers s.t. $x^i = \begin{pmatrix} x_1^i \\ x_2^i \end{pmatrix}$ is a $2-$dimensional vector representing index of line and index of column of center x^i.

Let also $D(x^k, x^i)$ be the Euclidean distance between x^k and x^i where i ranges from 1 to n and k is an random value chosen from 1 to n. $D(x^k, x^i)$ is computed as follows:

$$D(x^k, x^i) = \sqrt{(x_1^k - x_1^i)^2 + (x_2^1 - x_2^i)^2} \quad (1)$$

By sorting $D(x^k, x^i)$ and taking the first N points which satisfy this sorting, we will have a subset of X with N elements each representing a connected components the closest to x^k.

Le $Y = y^1, y^2, \ldots, y^N$ be the set of N closest connected components to x^k. Geometrically, this set can be seen as a polygon centered at x^k and having at most $N-1$ vertices; number of vertices can be less than $N-1$ because some vertices can be aligned. From the set Y, we would like to get the best match of x^k to one of y^i, i.e. we would like to know if x^k and y^i are aligned and how close they are. We do this by computing the inner angle between x^k, y^i and the origin x^0.

If x^k is the first element of X, then, x^0 has coordinates x_1^0 and $x_2^0 = x_2^k + 1$. This means that x^0 is aligned to x^k and y^{i*} (the best match to x^k) must minimize the inner angle with the horizon. We choose x^0 aligned to x^k because, for the first element, we suppose that the writing in the manuscript is horizontal. We treat the case where the writing takes another direction (say up to the right or down to the right) by taking the history of Y later on. We explain more specifically how we compute and minimize the angle between x^k and y^i as follow. We keep the same notation as previously and use (without justification) some basic notions of geometry.

The inner angle between x^k, x^0 and y^i centered at x^k can be derived as follows.

First, the vector between x^k and x^0 is computed using:

$$\overrightarrow{x^0x^k} = x^0 - x^k \quad (2)$$

$$\overrightarrow{x^0x^k} = \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix} - \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} \quad (3)$$

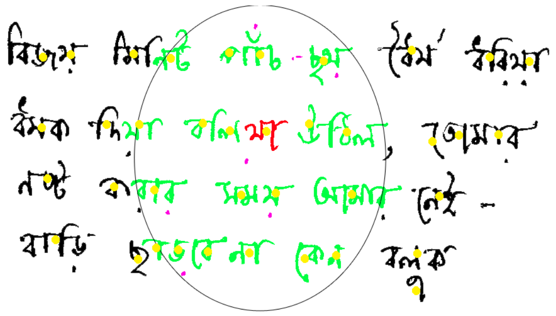


Figure 1: Example of connected component analysis for line segmentation. Red region: connected component to be matched. Gray circle: region of interest which is the set of connected components nearest to connected component to be matched (red region) centred at it. The region of interest include small connected components to be removed (pink regions) and connected component to be considered for comparisons and matching (blue regions). Black regions are connected components far from the connected component to be matched (red region) and are not considered for comparisons and matching. Yellow dots are centres of connected components. Note that centres of connected components are put only for visualization and are not effective centres returned by our implementation.

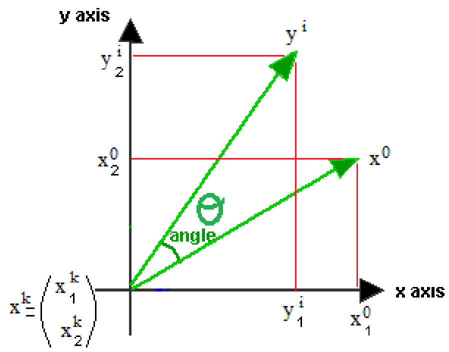


Figure 2: Example of $\widehat{x^0x^ky^i}$ computation; the inner angle between x^0, x^k and y^i centred at x^k.

Then, the vector between x^k and y^i is computed using:

$$\overrightarrow{x^ky^i} = x^k - y^i \tag{4}$$

$$\overrightarrow{x^ky^i} = \begin{pmatrix} x_1^k \\ x_2^k \end{pmatrix} - \begin{pmatrix} y_1^i \\ y_2^i \end{pmatrix} \tag{5}$$

The dot product of $\overrightarrow{x^0x^k}$ and $\overrightarrow{x^ky^i}$ is given by

$$\overrightarrow{x^0x^k}.\overrightarrow{x^ky^i} = \| \overrightarrow{x^0x^k} \| \| \overrightarrow{x^ky^i} \| cos\theta \tag{6}$$

where $\widehat{x^0x^ky^i}$ is the inner angle between $x^0x^ky^i$ cenered at x^k and $\| \overrightarrow{x^0x^k} \|$ is length of the vector $\overrightarrow{x^0x^k}$ and is computed as:

$$\| \overrightarrow{x^0x^k} \| = \sqrt{(x_1^0 - x_1^k)^2 + (x_2^0 - x_2^k)^2} \tag{7}$$

$\| \overrightarrow{x^ky^i} \|$ is computed the same way.
From equation 6, we can have the angle θ as follows:

$$\theta = arcos\left(\frac{\overrightarrow{x^0x^k}.\overrightarrow{x^ky^i}}{\| \overrightarrow{x^0x^k} \| \| \overrightarrow{x^ky^i} \|}\right) \tag{8}$$

Figure 2 shows a general case of computation of angle between two the three points ($\widehat{x^0x^ky^i}$). In the case of the figure, x^0 is not the horizontal axis but instead an arbitrary point in the space.

Once angles between x^k and all y^i computed, the next step is to choose the best match of x^k to one of y that minimizes the angle computed previously as follows: While x^k and x^0 are fixed for the point x^k, the only variable is y^i. Let then θ in equation 8 be equal to $f(y^i)$. Minimization of the angle ($\theta = f(y^i)$) is then subject to y^i and is performed as follows:

$$\underset{y^i \in Y}{\operatorname{argmin}} f(y^i) := \{y^* \| y^* \in Y \wedge \forall y^i \in Y : f(y^*) \leq f(y^i)\} \tag{9}$$

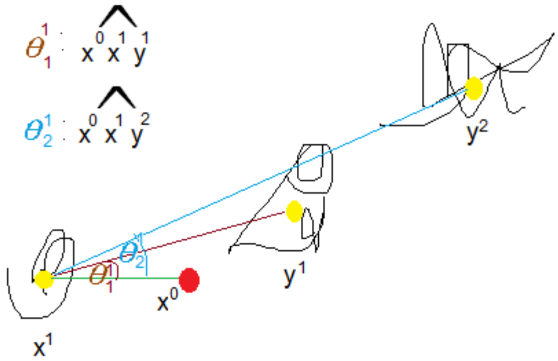


Figure 3: Example of angle minimization when x^k is the first connected component center in X. Black regions are connected components and yellow regions are centres of each connected component.

Figure 3 shows and example of this minimization. In the figure y^1 was chosen as the best match to x^1 since θ_1^1 (the angle $\widehat{x^0x^1y^1}$) was smaller than θ_2^1 (the angle $\widehat{x^0x^1y^2}$).

In a more general case, we choose y^* that minimizes the angle between the origin x^0 and current point to match x^k. A further refinement is done where if the angle constituted by x^k and y^* with the origin is higher than a certain threshold ε_θ , the matching is rejected. Note that, this threshold is initialized to $+\infty$. We update this ε_θ after we accept the first match.

Once the first match of x^k to y^* is chosen, update of ε_θ is necessary. ε_θ in this step will be the angle $\widehat{x^0x^ky^*}$). Update of threshold is necessary because if we accept the minimum angle without thresholding, then we can match a connected component in the end of the line to a connected component of another line. Thresholding allows stopping tracking of the connected component especially when the line in the manuscript ends.

At this point, the matching will result in a 3−dimensional vector where first dimension is x^k, second is y^* and third dimension is $\widehat{x^0x^ky^*}$.

When x^k is not the first point to be matched, then, to get the history of x^k we look at the first previous neighbour of x^k which is x^{k-1}.

The process starting from the choice of k in x^k to the angle minimization step is repeated to all points in X.

For remaining points in X, one obvious observation is that, remaining points will have a history, i.e. at least one previous point in X has already been matched. This is crucial to know the writing style in the manuscript. For example, if the writing in the manuscript was from going from up to down of the page, then, the origin x^0 will not be the horizontal line,

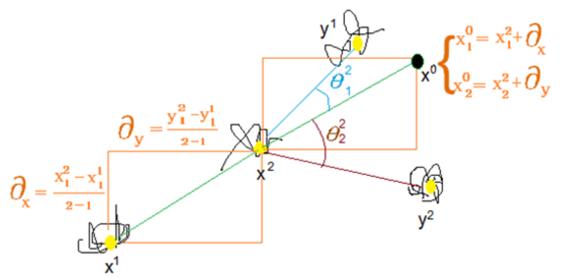


Figure 4: Example of angle minimization when x^k is not the first connected component center in X (angles have history). Black regions are connected components and yellow regions are centres of each connected component. The best match of $x^1$is given to y^1.

but will instead, will be calculated as follows:
Let x^k be the point to be matched and $x^k - 1$ be the first previous neighbour of x^k. To compute the hostory of writing i.e. the previous angle, we need to compute x^0 coordinate. coordinates of x^0 as computed as follows:

$$\partial x = \frac{x_1^k - x_1^{k-1}}{k-(k-1)} = x_1^k - x_1^{k-1} \tag{10}$$

$$\partial y = \frac{y_1^k - y_1^{k-1}}{k-(k-1)} = y_1^k - y_1^{k-1} \tag{11}$$

where ∂x and ∂x are displacement from x^{k-1} to x^k in the horizontal axis and vertical axis respectively.
x^0 coordinates can then be computed as follows:

$$\begin{cases} x_1^0 = x_1^k + \partial x \\ x_2^0 = x_2^k + \partial y \end{cases} \tag{12}$$

where x_1^0 and x_2^0 are the row and column indices of x^0 respectively.

This computation is important since we would like to keep history of writing which is based in the previous matching. This follows the principle of different types of tracking especially the Kalman filter [ref] which keeps history of previous observations. Note that our approach is different from tracking approaches as we do not keep a whole history but only the previous observation of angle which is most similar to Hidden Markov models approaches.

Once x^0 coordinates computed, the same process is repeated to get the best match of x^k. Figure 4 shows an example of x^0 computation and getting the best match according to the angle history. In the figure, the best match would be given to y^1 instead of y^2 although the angle between x^1, y^1 and the $x-axis$ is bigger than the one between x^1, y^2 and the $x-axis$, but since the matching follows same pattern as previous matching, the best match is given to y^1.

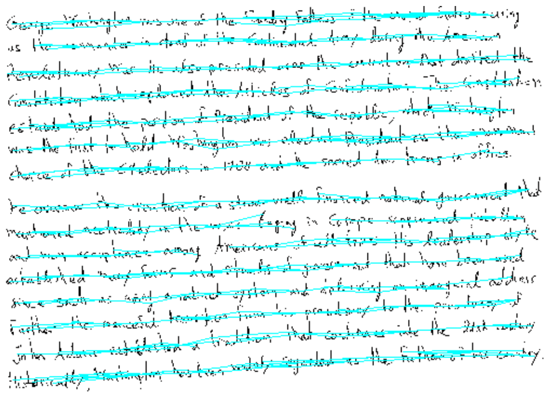
3 EXPERIMENTAL RESULTS

We tested our approach on images of ICDAR 2013 Handwriting Segmentation Contest, (Stamatopoulos et al., 2013). The dataset consists of 150 document images written in English and Greek as well as 50 images written in Bangla along with the associated ground truth for training and 50 images written in English, 50 images written in Greek and 50 images written in Bangla for test (Stamatopoulos et al., 2013). The dataset is challenging in that the skew angle between text lines and within the same line is different.

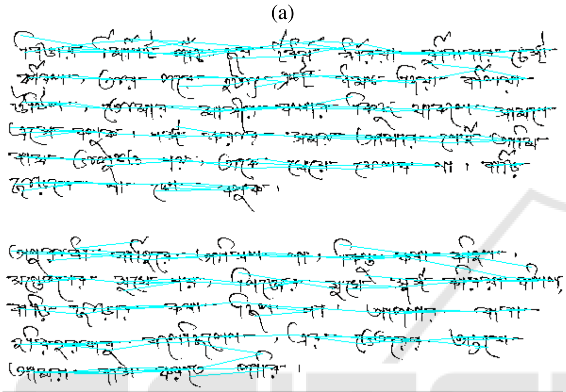
We implemented our approach in matlab 2010 and we choose in our approach the angle threshold $\varepsilon_\theta = 0.2^o$, number of neighbours of the connected component to track as $N = 20$ and minimum size of the connected components as 200 pixels.

In the experimental phase, we draw a blue line between each pair of centres of matched connected components found in section 2. When lines are cumulated, they show clearly direction of motion of words (connected components that we matched previously). Although visual observation can prove robustness of the approach, quantitative analysis is necessary to validate the approach and extend it to other datasets. We did not include the quantitative analysis in this work because our approach links only pairs of lines. If we apply the software proposed in the contest as described in section II and III of (Stamatopoulos et al., 2013) and in (con,) we would have a low accuracy. This is because each pair of connected components would be considered as a line. In order to solve this issue we propose two solutions: either to cluster blue lines (between matched connected components) so that each cluster constitute a line in the handwritten document, or to track, among all pairs, the first connected component so that to have a complete trajectory of it in the line. We leave this improvement to a future work while we present here only the main steps of the approach and a preliminary result.

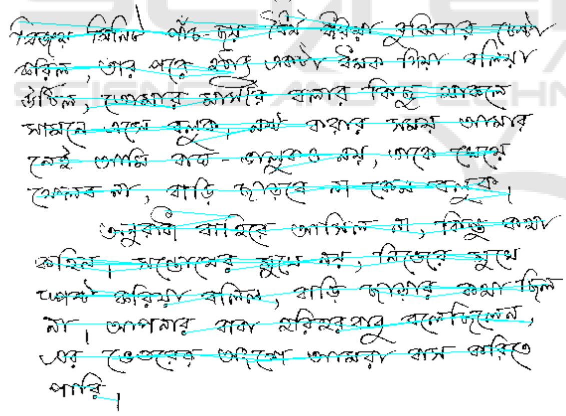
Figure 5 shows examples of application of our approach to images of the contest. Although images are in different orientations, our approach can still detect lines in the handwritten documents. However, two main drawbacks can be observed in the approach; first, several connected components were ignored in the processing, those components are the one smaller than the threshold size defined as 200 pixels previously. The second drawback can be observed in the two last handwritten lines of figure 5.(b). In the figure, we can observe some blue lines from the two handwritten lines merged. Since some components have their centres further from center of the word they belong to, and due to their previous observation, they



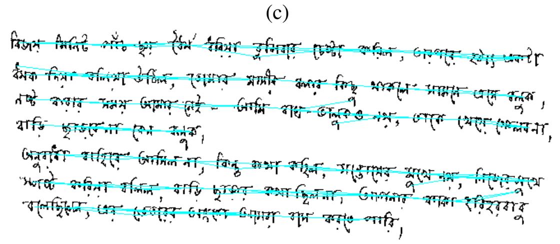
(a)



(b)



(c)



(d)

Figure 5: Result of our approach on images of ICDAR 2013 Handwriting Segmentation Contest. (a) result of processing image 202.tif, (b) result of processing image 337.tif,(c) result of processing image 342.tif,(d) result of processing image 343.tif.

are matched wrongly. One possible solution can be to enhance resolution of the original image so that each line keeps its own connected components close to each other.

4 CONCLUSION

In this work, we presented a new approach for handwritten text line segmentation inspired of various tracking approaches. The aim of the approach is to track each pair of connected components in the handwritten document which satisfy angle minimization. The approach is suitable when when connected components in the handwritten document are close to each other independently of the skew and line orientation. However, the approach can fail when connected components from different lines are close. Inthis case, they are merged to same line since the most available information used in our approach is the center of the connected component. The approach gave acceptable visual results but need to be enhanced with a complete tracking so that a quantitative analysis can be performed. The approach being innovative can be enhanced and open a new way of detecting lines in handwritten documents using word tracking.

ACKNOWLEDGEMENT

Authors would like to thank program chairs of the International Document Image Processing Summer school (IDIPS 2015) (IDI,) for their introduction to the ICDAR 2013 Handwriting Segmentation Contest during the summer school and for their pertinent explanations and remarks.

REFERENCES

http://users.iit.demokritos.gr/∼nstam/ICDAR2013HandSegmCont.

http://samosweb.aegean.gr/idips2015/.

Bai, N. N. (2008). Extracting curved text lines using the chain composition and the expanded grouping method.

Bar-Shalom, Y. (1987). *Tracking and Data Association*. Academic Press Professional, Inc., San Diego, CA, USA.

Boykov, Y. and Huttenlocher, D. P. (2000). Adaptive bayesian recognition in tracking rigid objects. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 697–704 vol.2.

Broida, T. J. and Chellappa, R. (1986). Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(1):90–99.

Bukhari, S. S., Shafait, F., and Breuel, T. M. (2008). Segmentation of curled textlines using active contours. In *Document Analysis Systems, 2008. DAS '08. The Eighth IAPR International Workshop on*, pages 270–277.

Gatos, B., Pratikakis, I., and Ntirogiannis, K. (2007). Segmentation based recovery of arbitrarily warped document images. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 989–993.

Goto, H. and Aso, H. (1999). Extracting curved text lines using local linearity of the text line. *International Journal on Document Analysis and Recognition*, 2(2):111–119.

Hönes, F. and Lichter, J. (1994). Layout extraction of mixed mode documents. *Machine Vision and Applications*, 7(4):237–246.

Li, X., Wang, K., Wang, W., and Li, Y. (2010). A multiple object tracking method using kalman filter. In *Information and Automation (ICIA), 2010 IEEE International Conference on*, pages 1862–1866.

Li, Y., Zheng, Y., Doermann, D., Jaeger, S., and Li, Y. (2008). Script-independent text line segmentation in freestyle handwritten documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1313–1329.

Loo, P. K. and Tan, C. L. (2002). *Document Analysis Systems V: 5th International Workshop, DAS 2002 Princeton, NJ, USA, August 19–21, 2002 Proceedings*, chapter Word and Sentence Extraction Using Irregular Pyramid, pages 307–318. Springer Berlin Heidelberg, Berlin, Heidelberg.

Louloudis, G., Gatos, B., and Halatsis, C. (2007). Text line detection in unconstrained handwritten documents using a block-based hough transform approach. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 02*, ICDAR '07, pages 599–603, Washington, DC, USA. IEEE Computer Society.

Mitiche, A. and Aggarwal, J. (2014). *Computer Vision Analysis of Image Motion by Variational Methodsn*. Springer International Publishing.

Pal, U. and Roy, P. P. (2004). Multioriented and curved text lines extraction from indian documents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(4):1676–1684.

Pal, U., Sinha, S., and Chaudhuri, B. B. (2003). *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29 – July 2, 2003 Proceedings*, chapter Multi-oriented English Text Line Identification, pages 1146–1153. Springer Berlin Heidelberg, Berlin, Heidelberg.

Roy, P. P., Pal, U., and Lladós, J. (2012). Text line extraction in graphical documents using background and foreground information. *International Journal on Document Analysis and Recognition (IJDAR)*, 15(3):227–241.

Stamatopoulos, N., Gatos, B., Louloudis, G., Pal, U., and Alaei, A. (2013). Icdar 2013 handwriting segmentation contest. In *2013 12th International Conference on Document Analysis and Recognition*, pages 1402–1406.

Zheng, B., Xu, X., Dai, Y., and Lu, Y. (2012). Object tracking algorithm based on combination of dynamic template matching and kalman filter. In *Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on*, volume 2, pages 136–139.