

Research of the Handwriting Recognition Methods

^I Alexey I. Martyshkin, ² Evgeniy I. Markin ^{1,2} Sub-Department of Computers and Systems, Penza State Technological University, 440039, Russia, Penza, Baydukov Passage / Gagarin Street, 1a/11 *alexey314@yandex.ru evgeniymarkin1@gmai.com*

Received: 21st August 2020, Accepted: 14th September 2020, Published: 31st October 2020

Abstract

This paper examines the methods for handwriting recognition. In the first part, the basic concept of the text recognition system is considered and the features of the operation of these systems are described, it also talks about possible problems in recognition that arise when working with text. The second part examines the difficulties in recognizing handwritten characters associated with the peculiarities of a language, style and handwriting recognition. The third part contains an analysis of methods for preliminary processing of images to improve the quality of further segmentation and text recognition. The next section describes approaches to segmenting text in a document. The fifth section contains an analysis of some of the methods for classifying handwritten characters. This section talks about the principles of such methods as using templates, using outline models, and also attribute-based methods. The mathematical models and algorithms for each method are described. The conclusion summarizes the state of the handwriting recognition systems and provides a general algorithm for the system's operation in order of its further implementation based on the above analyses.

Keywords

Handwriting Recognition, Handwriting Segmentation, Methods of Handwritten Character Classification, Attribute-based Classification Method, Method of Outline Models.

Introduction

Handwriting recognition (HWR) is a more difficult task than typed text recognition and has not been fully resolved at the moment [1].

Handwriting recognition can be divided into two types [2]:

- Online recognition is the recognition of text in the course of its entering into the system;
- Offline recognition is the recognition of text already entered into the system as an image.

Online recognition means handwriting recognition at the time of text writing. This approach allows the handwriting recognition system to obtain additional information about the speed of movement of the pen, writing pressure, the principle of writing a character, etc.

Only graphic information is available to the system in the task of offline recognition, which complicates the recognition process in comparison with online recognition.

At the moment, commercial HWR packages can confidently recognize only printed forms (questionnaires, completed forms, etc.), since when using documents of a strictly defined form and limiting possible input characters, the quality of the recognition task becomes much easier.

Figure 1 shows the general scheme of the offline handwriting recognition system [3].

At the first stage, the scanned image of a document, which is a bitmap image, is fed to the system input. Poor image quality can greatly affect the final result; for this, the input image is pre-processed to improve its quality before the start of recognition, since noise, blur, and low contrast complicate the task for recognition algorithms.



Figure 1: Scheme of the Offline Text Recognition System Operation

Next, the image obtained after the preprocessing procedure goes through the segmentation. In the course of segmentation, the basic units of text are determined, such as: lines, words and symbols. Obtaining fragments of higher levels (lines and words) can be carried out by analysing the gaps between the dark areas.

This approach is not suitable for identifying individual characters because when writing words, symbols are often bounded to each other and represent one connected element (Figure 2).

связанность

Figure 2: An Example of Elements Bounded Among Themselves

Also, the image of one letter can have the effect of decay due to poor scan quality or physical damage to the paper carrier, as shown in Figure 3. A human, unlike a program, can recognize that it is one single character. To solve the symbol segmentation problem, heuristic algorithms of increased complexity are often used. They consume a lot of computing power. At this stage of processing, the recognition system will not have enough information to make a decision about passing the border of a specific letter.

телефон

Figure 3: Decomposition of Letter Images into Unbound Components due to Poor Scan Quality

The result of segmentation is a segmentation tree, which is a data structure that displays the structure of text on an image. At the top level, there is an original image, which contains an array of string objects. Each string consists of a set of objects, which are words, which in turn are the leaves of the resulting tree.

Information about possible places where a word is divided into letters is stored in the word itself; separate objects for letters are not selected. Each tree object stores information about the area occupied by the corresponding object in the image.

The resulting image fragments are fed to the classifier's input, the result of which operation is the probability vector that the character belongs to the class of a particular letter.

Main Difficulties of Character Recognition

Character recognition is a very difficult task, in which there are many problems [4], the most significant of them are the following:

- A variety of languages, characters, and ways to represent them. Many languages have similar letters: "N" and "U", "R" and "Я", etc. In every language there are symbols that have similar features to each other and numbers. For example, in Russian they are: "M", "H", "W" and "O" and "O", etc.

- Distortion of character images, the reasons for which can be completely different, i.e. problems associated with the original image.

-A large number of variants of writing characters: by their size, slope, set of components, connecting elements between them, etc.;

- Spelling errors in the text;

- Specific features of the faces, which do not allow separating characters confidently;

- Intersection of text elements, imposition of parts of text on top of each other; soiled places, blots, corrections, paper defects, as well as artefacts that occur during scanning;

- Non parallel lines of text; other.

Distortions of digital images of text characters can be caused by the following reasons [5]: printing noise, spots and false dots on the background near characters, etc.; shifting characters or parts of characters relative to their expected position in the string; changing the slope of the characters; distortion of the character shape due to the digitization of the image with "coarse" discreteness; lighting effects (shadows, identifies, etc.) when shooting with the camera.

Many OCR systems perform best when processing black text on light backgrounds. In the case of text documents, this is not critical. But if you use such a system for processing non-standard images: scans of passports, driver's licenses, and car license plates, there is a problem with the inability to obtain correct data without preliminary image processing.

Many of the problems described above can be eliminated at the preprocessing stage [6].

Analysis of Image Processing Methods

Before the start of document segmentation, an image preprocessing takes place with the aim of: improving the image quality; it is performed by image processing methods (filtering, noise reduction, contrast enhancement, etc.); identifying an area of interest in an image; image analysis methods are used and are designed to get rid of non-textual information (for example, spots, blots, images in text, and others).

Improving Image Quality

At this stage, the image is cleaned from scanning defects. This is done using standard image processing methods. In particular, at the very beginning of work, a Gaussian filter is often applied to the image for noise reduction [7].

An important role in image processing is played by the so-called threshold binarization, which is the conversion of an image into black and white format from colour one, or in format with shades of grey. This allows separating text and background, simplifies the application of many algorithms in the future, and also eliminates some noise in the image.

As a result of binarization, two peaks will be located on the image brightness histogram:

- High level located in the area of light pixels, i.e. set by the background;

- Low level in dark areas, which denotes text.

As a result, the task of searching for symbol elements is reduced to finding threshold values of brightness; it is the task of finding the optimal value between two peaks of the histogram. The solution to this problem can be realized using the Otsu method [8] and its various variations.

Identifying the Region of Interest

At this stage, in the binarized image the areas are selected, in which the text for recognition is located, and all unnecessary elements that may complicate recognition are removed. Such elements include blots, spots on paper that were not removed during the binarization process, pictures, etc. To remove them, we can, for example, identify the connection between the elements in the image, determine geometric features, and classify the connections based on them as part of the text or a defect using the heuristic method.

Analysis of Segmentation and Text Normalization Methods

Text segmentation is the division of text into easy-to-analyse constituent parts such as strings, words and characters.

Two main approaches can be distinguished among the methods of text segmentation: the "bottom-up" approach, and the "top-down" approach. A bottom-up approach identifies connections in text and then combines them into larger elements such as words, lines, and sections. A top-down approach uses sequential division of the entire document into smaller parts.

Segmentation of Strings

The task of segmentation of strings in printed documents is now considered fully resolved [9]. And in the general case, upon solving the problem of handwriting recognition when dividing strings, difficulties arise that do not allow directly applying algorithms suitable for printed text [10]: strings may not be only parallel, but they may also bend; different strings may be too close, and elements of text belonging to different strings may overlap.

Among the methods of dividing strings in the text, one can distinguish grouping methods, the horizontal projection method, and the selection of baselines [11].

The horizontal projection method uses a top-down approach and is widely used in the problem of dividing strings in printed documents. The bottom line is the following: the horizontal projection profile is considered as a basic level for the image; this projection profile is considered the sum of all image pixels along the horizontal direction; then local minima are searched for on it. Since the work is carried out with a binary image, in which the pixels of the text have a value of 1, and the pixels of the background are 0, these minimums will correspond to the string spacing.

In the case of segmentation of handwriting with large string spacing, this approach may work, but, in general, this assumption is not met. However, it is possible to use modified methods based on the horizontal projection method, for example, using the local horizontal projection method [12].

Grouping methods are based on a bottom-up approach: connected elements are combined into strings, depending on their geometric properties, such as size, shape and distance between them [13].

In handwriting recognition, their use is limited due to the possible combination of parts of text from different strings into one connection. However, it is possible to use modifications of these methods, for example, using probability maps of the document instead of strictly connected components.

Base line identification is based on the idea that a person writes either "on" or "over" some imaginary line. This method tries to approximate this line and then reconstruct a string from it.

The paper [14] describes an approach based on the Hough transform. Since the Hough transform selects lines, it can be used to select base lines in case they are not too curved. In [15], the Hough transformation is applied to the centres of connected elements of text pixels. This approach requires strings of text to be close to straight lines, but it allows selecting strings located in an arbitrary place and going at arbitrary angles.

The method of local minima is described in [16]. The connectivity components of text pixels in it are considered as curved strings, at which local minimum points are sought. Then these minimum points are combined into straight lines.

The intersection between elements of different strings is not only a problem of string segmentation, but also of text recognition, since assigning an element to the wrong string obviously impairs its recognition. Overlapping components are a problem for horizontal projection methods and grouping methods.

To search for intersecting elements from different strings, we can use such characteristics as the size of connected text elements, the fact that one connection has been assigned to several strings, or, on the contrary, not assigned to any string. After finding such "dubious" connections, we need to determine whether they refer to a certain string, or whether they need to be decomposed into elements connected to different strings.

Segmentation of Words

At this stage of the recognition system operation, the selected strings of text are divided into separate words. Unlike printed text, in which the spacing between words is constant and the spacing between characters within a word is much smaller than the spacing between words, recognition the size of the spacing between words in handwriting can be completely different.

Typically, a bottom-up approach is used to solve the word segmentation problem. The connected elements of the text referred to one string at the previous stage of the system operation are divided into words at this stage.

The work [17] describes the use of the Murr-Hildreth algorithm for identifying words in an image. After applying this method, closely connected elements of the text begin to merge together; this is used to identify words in the original image.

Also, to solve the problem of identifying words, we can start from the question "is the space between the connected elements of the text a space between words or an interval between letters in a word" and solve this problem using classification methods. The distance between parts of one word is usually less than between words; it is common to use the so-called "gap metrics": a certain value is calculated for the interval between connected elements and compared with a threshold value.

Normalization

Due to the fact that there are a huge number of variants for spelling words, their recognition is a very difficult process. Normalization is carried out to bring a word to a certain general form, without losing useful information necessary for further recognition. Slope correction and slant correction are some of the most commonly used normalization methods.

Slope is the amount (angle) of word deviation from the horizontal string. Slant is the amount of deviation of those elements of the word that should be vertical, from, in fact, the vertical line [18]. The angle of inclination of a word does not affect its meaning, but may impair its recognition.

The simplest slope correction method is as follows: to introduce a quality functional on the horizontal projection profile of the word image (for example, the standard deviation) depending on the image rotation angle; to maximize it over a range of angles.

In fact, the method repeats the method of "horizontalizing" the image in the OCR task. The principle of operation can be seen in Figure 4.

In [19], a correction method using the slope method for identifying the central region of a word is proposed. The central region of a word is the region that contains the major portions of the uppercase letters of the word.



Figure 4: Slope Correction using a Horizontal Projection Profile

A horizontal projection profile analysis is used to find the central region. Namely, the global maximum is found on the projection profile, and its edges are taken. Further, the baseline of the word is approximated by the minima of the strokes lying inside the central region. The image is then rotated so that the baseline is horizontal. In [20], a modification of the method for identifying the central region by analysing the distribution of the values of the horizontal projection profile is proposed.

There are other methods, for example, based on smoothing and linear regression (Figure 5).

There are other methods of normalization, such as resizing and identifying the skeleton of the text, but they are used less often.

Helix (2020) 10 (5): 62-68



Figure 5: Slope Correction using RLSA and Linear Regression

Analysis of Character Classification Methods

Character classification is the main task of image character recognition; each individual character is taken and classified at this stage. There are several methods for classification, the most common are: template-based methods; methods using outline models; attribute-based methods; structural spot reference methods; hidden Markov model; k-nearest neighbours method; method based on the skeleton structure of a character.

Methods using Templates

Today the most common template method is the method for calculating the correlation coefficient between two matrices.

In addition to the original image, the method based on the correlation coefficient uses reference images being images of the recognized object [21].

The image is displayed in the computer memory in the form of a matrix. For example, A is the matrix of the classified image, and B is the reference; the sizes of the matrices are $m \times n$; the correlation coefficient is calculated by the formula 1.

$$k = \frac{\sum_{m} \sum_{n} (A_{mn} - \overline{A})(B_{mn} - \overline{B})}{\sum_{m} \sum_{n} (A_{mn} - \overline{A})^2 (B_{mn} - \overline{B})^2},\tag{1}$$

Where \overline{A} , \overline{B} are mean values of matrix elements.

To calculate the correlation value, we need to scale the matrices to equal sizes; therefore, a matrix, which will be equal to the size of the reference matrix, is cut from the matrix of the original image.

The method based on calculating the correlation coefficient is common, despite the fact that it is very expensive in terms of computational complexity. This is due to the need to scale and rotate the original image.

Using outline Models

One of the well-known methods for identifying outlines is the spatial differentiation method. It is based on the estimation of the rate of change (gradient) in the brightness of each original image point. A point belongs to the outline if it is on the border of two regions of different brightness, i.e. brightness changes quickly enough [22].

Spatial differentiation operators are high cut filters. The most common operator masks are the Sobel, Prewitt, or Roberts operators (Figure 6), which, for the case of identifying horizontal drops (calculations of $g_y(i,j)$) have the following form.

Sobel				Prewitt				Roberts			
[-1	-2	-1	1	[-1	-1	$^{-1}$	1	[O	0	0	1
0	0	0	1	0	0	0	:	0	1	0	0
11	2	1	1	l 1	1	1	1	lo	0	-1	1

Figure 6: Masks of Sobel, Prewitt and Roberts Operators

The Sobel operator is a discrete differential operator that calculates an approximate value for the brightness gradient of an image.

The masks of the Sobel and Prewitt operators to identify the vertical changes in brightness are obtained from the masks of the same name to identify the horizontal drops by transposition, and the mask of the Roberts operator is obtained by rotating 90° clockwise.

The resulting array g(i, j) is called a gradient image. After that, the threshold processing of the gradient image is carried out according to rule 2.

$$b(i,j) = \begin{cases} 1, \text{at } g(i,j) \ge T \\ 0, \text{at } g(i,j) < T \end{cases}$$
(2)

Where T is a constant or variable threshold depending on the local properties of the image.

The result is a binarized (consisting only of black or white pixels) image b(i, j), the points of which correspond to the boundary points of the original image.

The main reason for the high popularity of these methods based on the use of spatial differentiation masks is the simplicity of their implementation on a computer.

Attribute-based Methods

Attribute-based methods are the methods based on the fact that an N-dimensional attribute vector is calculated for an image. The recognition process is based on the principle of comparing this vector with a set of reference vectors of equal dimension. Recognition and decision-making on assigning a certain class to a character occurs based on the analysis of calculated attributes and has a number of rigorous mathematical solutions within the framework of the probabilistic approach [23]. Most often, a classification is used in character recognition systems, which is based on calculating the Euclidean distance between the attribute vectors of the recognized character and the attributes of the reference description. The compilation of the attribute vector is carried out in the process of image analysis. This process is called attribute extraction.

The main advantages of attribute-based methods are: ease of implementation; high resistance to changes in the shape of characters; low number of refusals from recognition; high performance.

The disadvantages include: instability to various image defects; irreversible loss of part of the information about the character that occurs at the stage of attribute extraction; the extraction of attributes is performed independently; therefore, information on the mutual arrangement of character elements is lost.

E.S. Abramov suggested using the m_k sequences as image attributes, which are calculated as follows [24].

Let the image A have an object O. Take N radius of r_k vectors with the beginning in the centre of the image C and the end at the border of the image, and standing from each other at an angular displacement of $2\pi/N$. Along each vector, there are image points belonging to the object O, and points not belonging to this object. These points form segments B_{kl} , $1 \le l \le N_k$, which represent parts of the object O along the vector. The length L_{kl} and its distance from the center of the image d_{kl} are determined for each of these segments. The quantity determined by the expression $m_r = \sum_{l=1}^{N_k} d_{kl} L_{kl}$ is called the moment of the image along the vector r_k . After normalization $m_k = \frac{m_k}{\sum_k m_k}$, the invariance of the image moment to scaling is ensured, which makes it possible to ensure the recognition

of objects in images regardless of their scale.

D.V. Gorlov proposed the idea of forming the attributes of the recognition system based on the function invariant to shift and rotation [25]. To do this, he calculated such parameters as height, width, number of points, and attributes for each character.

The attribute calculation procedure is divided into five stages:

1. The centre of gravity value is calculated for the outline of the current character.

For the current point of the circle u with the centre at the centre of its gravity and radius r = const, the value is calculated by the formula 3.

(3)

(4)

$$\mathbf{F}_Q = P_Q(\theta^2(x, u))$$

Where *x* is the current point of the character outline;

Q — number of attributes;

 P_0 — Legendre polynomial;

 $\theta(x, u) = p(x, u)/r; \eta = p^2(x, u); \eta$ — Euclidean distance on a plane.

For all points of the character outline, the values $F'_{Q-1} \dots F'_1$ are calculated.

For all points of the circle, there is a characteristic point at which the value of F'_{0} is maximum.

For this point of the circle, the values $F'_{Q-1} \dots F'_1$ are calculated.

Values of the form are used as attributes:

$$\mathbf{F}_Q = \frac{1}{k} \sum_{i=1}^{k} P_Q(\theta^2(x_i, u))$$

Where k is a number of points of the character outline.

The 1 / k multiplier was introduced to reduce the effect of image discreteness. Orthogonal Legendre polynomials are chosen as attributes:

$$P_1(\eta) = \frac{2}{b-a}(\eta - a) - 1, P_2(\eta) = \frac{1}{2}(3P_1^2 - 1), P_3(\eta) = \frac{1}{2}(5P_1^3 - 3P_1)$$
(5)

The classification is carried out as follows: parameters and characteristics are calculated for the character outline, and the character outline is subsequently compared with the objects obtained during machine training. In this case, the proximity measure is calculated in the form:

$$L = \sum_{i=1}^{Q} |\mathbf{F}''_{i} - \mathbf{F}_{i}|$$
(6)

Where F''_i is an object attribute obtained during training;

 F_i — object attribute obtained by recognition; Q — number of attributes.

Conclusion

In this paper, research was carried out on the difficulties of handwriting recognition and the main its stages, namely: image preprocessing, text segmentation and classification. At the stage of analysing the preliminary processing of the image, it was revealed that to obtain the maximum amount of useful information from the image of the text, it is enough to carry out the following actions: convert the image to black and white format; to binarize the image; apply the Sobel filter to the resulting image.

After that, the text is segmented; because text recognition will be performed in a given form, then it is enough to segment individual characters. For this purpose, an image obtained after applying the Sobel filter is used, from which information about the location of objects is obtained, containing necessary information. Based on this, individual characters are cut from the original image for further classification.

References

[1] Too J., Gonzalez R.: Principles of pattern recognition. Mir, Russia, Moscow (1978)

[2] Mansi Shah G. A literature review on hand written character recognition. Indian Streams Research Journal (2013)

[3] Dmitrievich K.V.: Algorithms for offline recognition of handwritten digits. Moscow State University named after M.V. Lomonosov, Moscow (2013)

[4] Bedi R.K.: Various image enhancement techniques: a critical review. International Journal of Advanced Research in Computer and Communication Engineering (2013)

[5] V.P. Kvasnikov: Improving the visual quality of a digital image by element-wise transformation. Aerospace engineering and technology (8), 200-204 (2009)

[6] Li Y.: Script independent text line segmentation in freestyle handwritten documents. Pattern Analysis and Machine Intelligence (8),1313-1329 (2008)

[7] Shapiro L.A. Computer vision. vol. 8. Knowledge laboratory (2006)

[8] Otsu N.: A threshold selection method from gray-level histograms. Automatica (11) 285-296 (1975)

[9] Plamondon R.A.: Online and off-string handwriting recognition: a comprehensive survey. Pattern Analysis and Machine Intelligence. vol. 1 (22), 63-84 (2000)

[10] Adamek T.A.: Word matching using single closed outlines for indexing handwritten historical documents. International Journal of Document Analysis and Recognition (IJDAR) (9) 153-165 (2007)

[11] Likforman-Sulem L.A.: Text string segmentation of historical documents: a survey. International Journal of Document Analysis and Recognition (IJDAR) (9) 123-138 (2007)

[12] Bar-Yosef I.A.: String segmentation for degraded handwritten historical documents. Document Analysis and Recognition, 1161-1165 (2009)

[13] Likforman-Sulem L.A. A Hough based algorithm for extracting text strings in handwritten documents. Document Analysis and Recognition (2), 774-777 (1995)

[14] Kim G.A.: Architecture for handwriting recognition systems. International Journal on Document Analysis and Recognition, vol. 1 (2) 37-44 (1999)

[15] Manmatha R.A.S.N.: Scale space technique for word segmentation in handwritten documents. Scale-Space Theories in Computer Vision 22-33 (1999)

[16] Vinciarelli A.A.L.J.: A new normalization technique for cursive handwritten words. Pattern Recognition Letters, vol. 9 (22) 1043-1050 (2001)

[17] Bozinovic R.M.A.S.S.N.: Off-string cursive script word recognition. Pattern Analysis and Machine Intelligence, vol. 1 (11) 68-83 (1989)

[18] Vapnik V.N .: The theory of pattern recognition. Publishing house Nauka-e M. (1974)

[19] Mestetsky L.M.: Mathematical methods of pattern recognition (course of lectures), Moscow State University named after M.V. Lomonosov (2004)

[20] Duda R .: Pattern Recognition and Scene Analysis, Mir, Moscow (1989)

[21] Laros D.T .: Gaining Knowledge in Data: An Introduction to Data Mining (2012)

[22] Bagrova I.A.: The choice of attributes for the recognition of printed Cyrillic characters. Bulletin of Tver State University (28) 59-73 (2010)

[23] Osetrova O.V.: Semiotics of a font. Voronezh State University Bulletin (2006)

[24] Kam H.T.: Proceedings of the 3rd International Conference on Document Analysis and Recognition. Random decision forests, Montreal (2007)

[25] Mestetsky L.M.: Continuous morphology of binary images of text in ABBYY in a modern way, FIZMATLIT, Moscow (2009)