Development of OCR system for portable passport and visa reader

Yury V. Visilter, Sergei Yu. Zheltov, Anton A. Lukin

ABSTRACT

The modern passport and visa documents include special machine-readable zones satisfied the ICAO standards. This allows to develop the special passport and visa automatic readers. However, there are some special problems in such OCR systems: low resolution of character images captured by CCD-camera (down to 150 dpi), essential shifts and slopes (up to 10 degrees), rich paper texture under the character symbols, non-homogeneous illumination. This paper presents the structure and some special aspects of OCR system for portable passport and visa reader. In our approach the binarization procedure is performed after the segmentation step, and it is applied to the each character site separately. Character recognition procedure uses the structural information of machine-readable zone. Special algorithms are developed for machine-readable zone extraction and character segmentation.

Keywords: optical character recognition, image segmentation, passport and visa documents.

1. INTRODUCTION

The modern passport and visa documents include special machine-readable zones satisfied the ICAO standards (fig.1,2). This allows to develop the special passport and visa automatic readers. A number of different firms and organisations (like the travel agencies) may use the corresponding OCR-systems (OCRS) as well as the official state departments.

It seems that the most popular and useful implementation of such OCRS will be the desktop version based on PC and desktop optical scanner. But in some applications users may need to operate with an autonomous portable passport and visa reader based on CCD-camera. This extended statement of problem makes very actual some special requirements that an improper for the ordinary general-purpose OCR-systems:

- the system has to recognise the characters at extremely low resolution (down to 150 dpi);
- the essential shifts and slopes (up to 10 degrees) of the documents in the camera frame are available and must be ignored at the stage of segmentation of machine-readable zone;
- the rich texture presents on the visa document right under the symbols;
- the non-homogeneous illumination and slight page deformations may occur in the case of image capturing by CCD-camera;
- the special structure of machine-readable zone has to be taken in account to provide the information retrieval;
- the high importance of the information (that is the identification of persons and their missions) forces the high requirements on the recognition accuracy (the probability of error character recognition 1.0E-04 - 1.0E-05).

This paper presents the research that was performed in order to develop the OCR-technology that will satisfy all these requirements. The following aspects of this work are outlined below:

a) modular structure of the algorithm;

b) document structure and the segmentation of machine-readable zone;

c) rows segmentation (determination of character sites);

d) character binarization;

e) character recognition;

f) information retrieval.

The special structure of the algorithm was developed. The original algorithms were developed for image and rows' segmentation. For other tasks the standard algorithms were adopted. Adaptation of these algorithms is not discussed here in details. The problems of hardware implementation and Russian characters decoding are out of discussion in this paper.

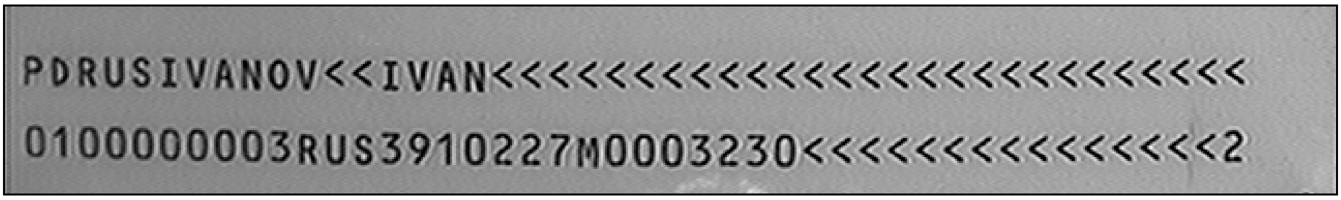

Fig.1. Example of passport machine-readable zone content.

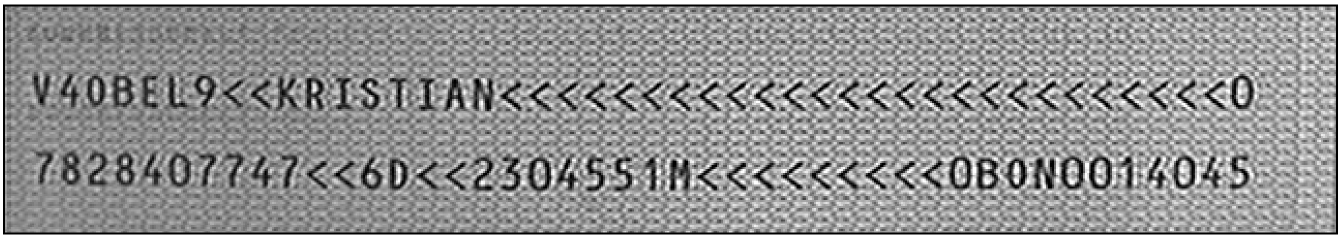

Fig.2. Example of visa machine-readable zone content.

2. MODULAR STRUCTURE OF THE ALGORITHM

All the modern OCR systems have a modular structure. They are built of the practically independent procedures and the system calls these procedures sequentially. The usual order of procedural steps is the following:

1) document binarization;
2) document segmentation (determination of rows and columns);
3) rows segmentation (determination of character sites);
4) character recognition;
5) information retrieval.

Moreover, often the image is captured as binary and the processing starts from the step 2.

The modular structure of our technology is a little bit different:

1) document segmentation (machine-readable zone extraction);
2) rows segmentation (determination of character sites);
3) character (on-site) binarization;
4) checking of structural constraints;
5) character recognition;
6) information retrieval by check digits.

The first difference is that the binarization is performed after the segmentation and it is applied to the each character site separately (with the proper adaptive threshold). Our reason here is that the non-homogeneous background (texture) and possible non-homogeneous illumination makes the analysis of grey level picture more appropriate for segmentation procedures (step 1-2). If we use the adaptive thresholding for each character site separately, the quality of characters binary images will be essentially better. It is especially important in the case of low resolution images. When the character size is 10x15 pels, the each pixel includes the crucial information.

The second original element in our scheme is a checking of structural constraints (step 4). The ISO standard defines a very complex and sophisticated structure for a machine-readable zone of passport and visa documents. This structure is designed in order to increase the recognition accuracy. It provides two possible ways to this purpose. Firstly, we know that some positions of the row may contain only the digits and in other positions - only the letters. Moreover, some positions allow the closer sets of characters. For example, the set {'V','P'} in the first position consists of two possible characters only. We refer this type of information as "structural constraints". Secondly, the check digits (control sums) control some groups of characters in this structure.

Theoretically, we can use the both types of structural information at the final stage, after the character recognition. But only if our recognition algorithm generates the belief measure for all hypotheses in a common set. Then we may choose the most likelihood decision from any given subset. However, some recognition algorithms do not this. The second reason to check the structural constraints before the recognition of characters is that some recognition algorithms directly compare the viewed characters or their descriptions and the model characters or model descriptions. These algorithms are better and faster since the set of models is closer to optimal one. So, at the step 4 we use the structural constraints. And at the step 6 we retrieve the information by the check digits.

In the next section the blocks of our processing scheme are briefly described.

3. ALGORITHM DESCRIPTION

3.1. Extraction of machine-readable zone

Machine-readable zone in the passport/visa documents consists of two rows of symbols in OCR-B code. The each row contains 44 characters. The vertical character size, the distance between the rows and the size of "clear fields" at the top and at the bottom of the zone are standard and known. The zone may be at the top or at the bottom of the image. But the exact position, orientation and slope of the zone in the field of camera view are unknown.

The procedure developed is modular too and consists of two steps:

1) rough localisation of machine-readable zone;

2) estimation of exact location and slope parameters.

The "rough localisation" is a fast step. At this step the narrow vertical stripe at the center of the image is analysed. We pick up two histograms of this stripe: intensity histogram and gradient magnitude histogram. In result we obtain the approximate knowledge about the location of machine-readable zone (plus/minus fifty pixels in height without the slope information).

For exact estimation of zone geometrical parameters we have developed the special voting algorithm based on ideas of Hough Transform[1-2]. The original Hough Transform is a well-known technique for straight line detection in the special parameter space (accumulator). Our approach also uses the Hough parameter space. But we use it in order to detect two groups of lines (two sloped stripes) that are corresponded to the rows of machine-readable zone (fig.3.). Tests demonstrate that the accuracy of this algorithm is about one pixel in height and about 1-2 degrees in slope angle.

Since the parameters of the zone are found, the algorithm extracts the machine-readable rows and transforms them into the horizontal view. As a result we obtain two images that contain the machine-readable rows only.

3.2. Rows segmentation

The images of the rows are processed separately.

Two lateral histograms[3] (in x-direction) are picked up at this stage. For the each pixel column the average intensity value and the difference between maximum and minimum values are calculated. Local maximal points in intensity histogram and local minimal points in difference histogram correspond to separators between the characters in the row (fig.4.).

Fusion of information from these two histograms makes it possible to provide the robust separation of character sites even under condition of non-homogeneous illumination of the document.

3.3. Character binarization

The purpose of this step is to obtain the binary images of characters. It is achieved by means of histogram-based adaptive thresholding. The thresholds are local and determined separately for each character site.

At the earlier stage of our work we used the well-known Otsu algorithm for threshold determination[4]. But it is not very fast and being applied 88 times (for each character in the zone) makes the whole program too slow. Now we use the analogous but more fast iteration version of this algorithm.

The morphological filtering[5] is applied to the results of binarization procedure in order to avoid the false points, “peaks” and “holes” of the character binary images.

3.4. Character recognition

Based on the reasons outlined in section 2, we have implemented and tested only those recognition techniques that directly compare the description of viewed characters relative to the given set of model descriptions, and form the likelihood (belief) estimates for all hypotheses from this set. Two methods were better then other we tried: the rough correlation technique, so called “zoning”, and the “crossing by lines” technique.

“Zoning’ in our implementation describes the symbol as its rough image of size 5x5 (called “zoning vector”). On the real image, the each pixel of this rough picture corresponds to the rectangular “zone” with sub-pixel boundaries. The elements of “zoning vector” contain the expected average values of corresponded “zones”. These values were obtained by the training on a set of samples. The belief of the hypothesis is estimated as a normalised correlation between the “zoning vector” of viewed binary image and the “zoning vector” of this hypothesis.

The classic “crossing by lines” takes into account only the number and the order of symbol intersections with straight lines from a special set. We have developed the new modification of this algorithm. We introduce the separation of crossing segments into two groups: “short” and “long” segments. The decision rule accounts them with different weights. So, this algorithm has three additional tuning parameters: the segment length threshold and the weights of “short” and “long” segments. The training procedure for parameter estimation is developed.

These two methods were tested in comparison to each other. “Zoning” demonstrates better results but just a little bit and not for all types of character printing quality. The main conclusion here is that the both techniques strongly depended on a quality of symbol binarization. Especially, in the case of low resolution images.

3.5. Retrieval by check digits.

The algorithm of passport and visa machine-readable data retrieval is strongly determined in ICAO Doc9303 (part 2). This algorithm was implemented and tested in our system.

The main conclusion here is that the current structure of machine-readable zone provides better retrieval support for visa documents than for passports.

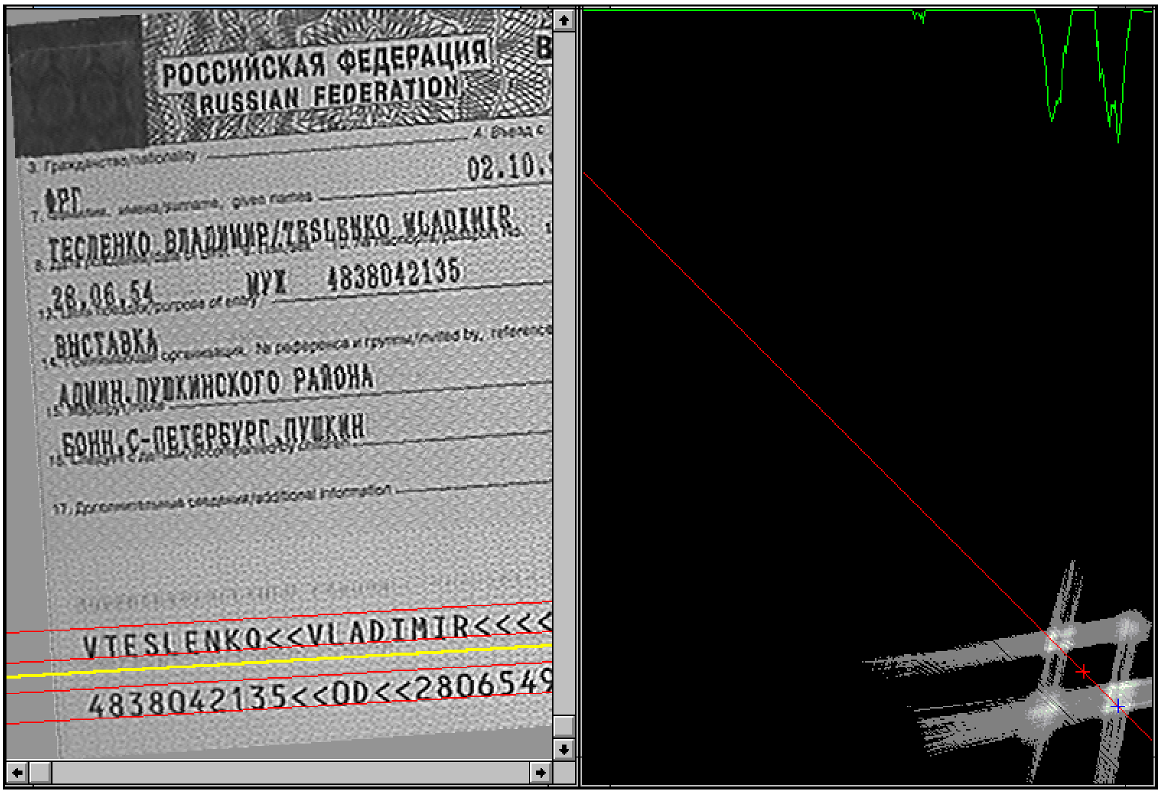

Fig.3. Machine-readable zone extraction stage. The Hough accumulator with natural parametrisation of straight lines is shown. Two machine-readable strings are found as a pair of “meander”-signals in corresponding accumulator section.

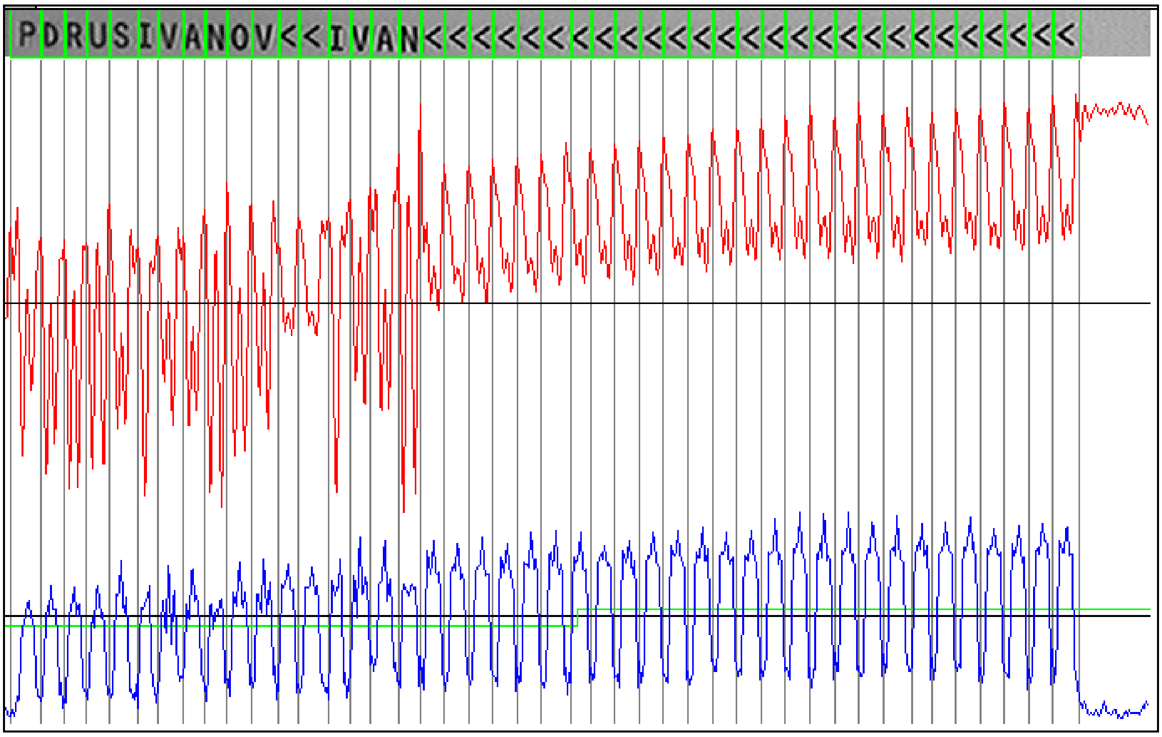

Fig.4. Row segmentation stage. Two lateral histograms are shown. Separation lines between characters correspond to maxima of intensity histogram and minima of difference histogram.

4. TESTS AND RESULTS

The algorithms proposed were implemented and tested on the wide set of real document images. The images were captured in different wave bands, by different scanners and CCD-cameras, with different resolution in the range 150-400 dpi. The detection and extraction of machine-readable zone were successful in 99.6% of all tests. The probabilities of character recognition (for “zoning”): at 150 dpi - 0.967; at 400 dpi - 0.992.

The exploration of recognition accuracy was performed in comparison to the popular in Russia commercial OCR-system “FineReader Professional 3.0”. The main result is that our system in the worst cases makes approximately half of errors relative to the “FineReader”. In our opinion, this effect is connected both with data retrieval possibilities of machine readable documents and with the other original features of our approach described above.

REFERENCES

1. Duda, R.O., and Hart, P.E. “Use of the Hough Transformation to Detect Lines and Curves in Pictures.” Comm. ACM, vol. 15, no. 1, pp. 11-15, 1972.
2. Nagao M., Nakajima S. "On the relation betweeh the Hough transformation and tne projection curves of a rectangular window", Pattern Recognition Letters 6(1987), 185-188.
3. Davies R. "Lateral Histograms for efficient object location: Speed versus ambiguity", Pattern Recognition Letters 6(1987), 189-198.
4. Otsu N. A theshold selection method from gray-level histograms //IEEE Trans. on System, Man and Cybernetics - 1979, v.SMC-9, N 1.
5. Giardina, C.R., and Dougherty, E.R. Morphological Methods in Image and Signal Processing, Prentice-Hall, Englewood Cliffs, N.J., 1988.