SPACE SYMBOL DETECTION ON COMPLEX BACKGROUND USING VISUAL CONTEXT

Alexander Sheshkus
Smart Engines Limited
Sham Peng Tong Plaza 103
Victoria, SEYCHELLES
E-mail: asheshkus@smartengines.biz

Vladimir Lvovich Arlazarov
Institute for Systems Analysis of
Russian Academy of Sciences (RAS)
117312, Moscow, Russia
Email: arl@isa.ru

KEYWORDS

Artificial neural networks, OCR, text line segmentation, monospaced font recognition.

ABSTRACT

In this paper, we present a new technique of space symbol detection in the task of monospaced text recognition for credit cardholder name as an example. In the considered case, standard methods fail or have low quality because of background complexity and variability of symbol colors. Suggested method is based on the usage of two conjoined symbol images as an input for artificial neural network. This provides background visual context for the recognizer, which helps it to distinguish between symbols and spaces.

INTRODUCTION

Warren McCulloch and Walter Pitt created the first computational model for artificial neural networks (ANN or simply NN) in 1943. They called this model threshold logic (McCulloch et al., 1943). The first known implementation of this model (Frank Rosenblatt, 1957) was designed for optical character recognition. Since 1980, usage of neural networks with convolution layers took place for the same task (Fukushima and Kunihiko, 1980) along with multilayer perceptron. However, computation efficiency of the previous generation of NNs is still attractive and people continue to develop new feature extraction methods for them (Anil K Jain et al., 1996).

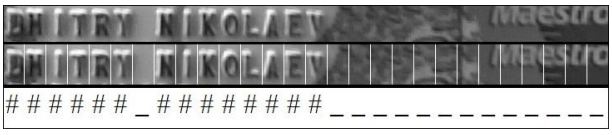
It is possible to write dozens of pages about symbol recognition with neural networks, but this work is about the application of NNs for something other than the recognition of separate symbols.

In this work, we consider one of the stages of cardholder name recognition on credit card images. We analyse the working high-performance card recognition system. One of the essential features of the system is processing time. Therefore, we could not afford methods with high computational complexity.

The first stage of the card recognition algorithm is card border detection. We use the method for rectangular document detection described in (Natalya Skoryukina et al., 2014). After obtaining card quadrangle, we use projective rectification to restore its original rectangular form.

Figures 1. Example of an Input Image for Card Recognition

On the rectified card image, we perform field detection and recognition. For field detection, we compute image projection on the vertical axis in the specified region of interest and determine top and bottom boundaries for every target field. In this paper, we consider only name field recognition. There is an example of the name field region of the input image on Fig. 2.



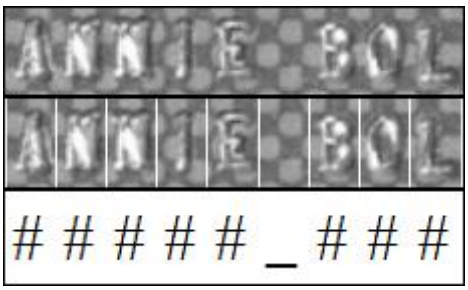
Figures 2. Example of an Input Image for Name Recognition, Character Boxes (“#”) and Space Positions (“_”)

On the name field region of the input image, we apply a segmentation method to find exact cell coordinates for every character. We use the fact that all symbols of the card name field originally have the same width, which noticeably simplifies the task. In such cases segmentation algorithm could be based on dynamic programming with character width deviation constraints. Algorithm itself is simple and works almost perfect for the analyzed system.

We crop symbol images and pass them as an input for the symbol recognizer based on an artificial neural

Proceedings 29th European Conference on Modelling and Simulation ©ECMS Valeri M. Mladenov, Petia Georgieva, Grisha Spasov, Galidiya Petrova (Editors)
ISBN: 978-0-9932440-0-1 / ISBN: 978-0-9932440-1-8 (CD)

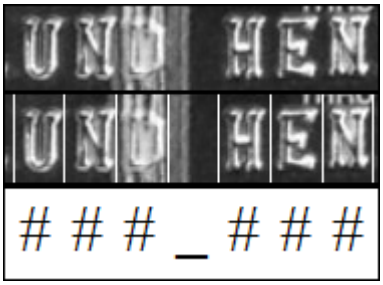
network. At the last stage of the name field recognition, we apply statistical post-processing using dictionaries (Sholomov et al., 2005; Slavin et al., 2011).



Figures 3. Image, which is Hard to Segment

Space/symbol detection precision of the recognizer used in our system is 96.5%, when its character recognition precision is 98.7%. The used recognizer has no special abilities for space/symbol detection and recognizes them as characters. Space is one of the 29 different symbols of the alphabet (A-Z, dot, space and hyphen) and has no special properties. As the recognition quality requirements increase, it becomes clear that dictionary based language model is also not powerful enough for correction of wrongly recognized spaces. Thus another algorithm for space/symbol detection is needed which would be efficient, robust and fast. This algorithm will have to deal with problems related to the complexity of card background: lines, curves, sparkles and shadows.

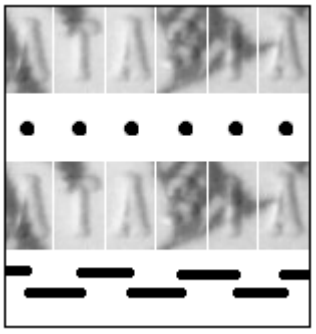
Standard methods of distinguishing between spaces and characters based on analysis of image statistics show unsatisfactory results (as we will show below) due to complexity of the background mentioned above. Due to the presence of sparkles and shadows along the strokes of the embossed characters the binarization methods, which are widely used in scanned document recognition (Jagroop Kaur and Dr. Rajiv Mahajan, 2014), are also inapplicable.



Figures 4. Non-Binarizable Image

In the considered case, any single character recognition algorithm may give very confident non-space answers on space images and very confident space answers on letter images, since even human not always able to tell for sure whether there is a letter or a space given cropped character cell image. On the other hand, space detection precision could be higher if the system will use a broader visual context (at least two neighboring characters).

We will use the NN with similar architecture, but in a different way. The main source of architecture difference between these neural networks is the processing time handicap. Therefore, new NN should be lighter. Instead of giving separated symbols for recognition, we will feed two neighboring symbols image as input (see Fig. 5). This NN will have four dimensioned output vector. This neural network is trained to separate two spaces, symbols and space, space and symbol and two symbols images.



Figures 5. Difference Between NNs Application: Base and Proposed Way.

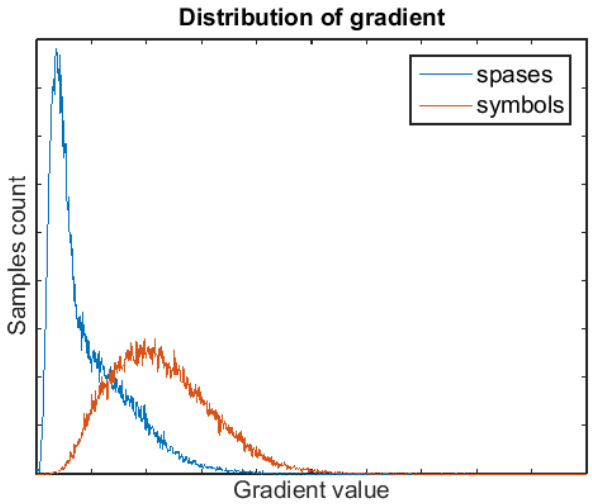
Now this approach is poorly investigated. Most relevant works consider handwriting letters recognition problem using ligatures classification (Bong-kee Sin and Jin H. Kim, 1997).

WHY NEURAL NETWORK?

Artificial neural networks are well-known technique (Tao Wang et al., 2012; Vivek Shrivastava and Navdeep Sharma, 2012), but it is not the only way to recognize symbols. A set of completely different methods also exists (Datong Chen et al., 2014), but every method has its own specifics and limitations. At first, let us consider simpler methods to understand why we want to use neural network in the task of space detection. The common approach to this problem is to collect and analyze symbol image statistics (Katherine L. et al., 2011). We assume that empty character box will differ a lot from the one with symbol in terms of gradient, dispersion, etc.

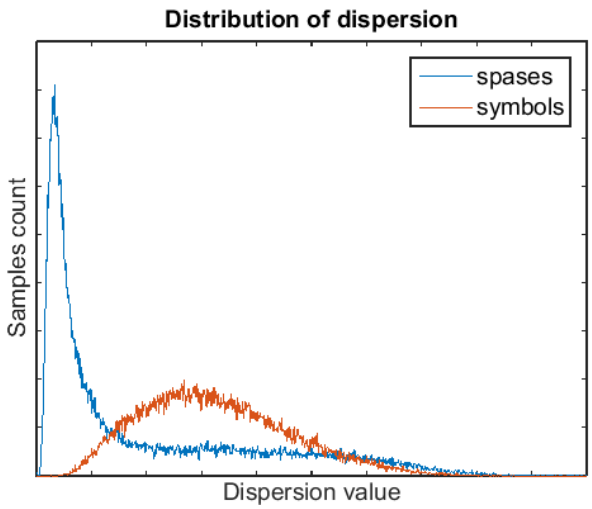
Let us consider in detail why these standard methods based on image statistics are inapplicable in our case. As mentioned above, the main problem is background complexity. We have analyzed some image statistics distributions to show a problem existence.

Trivial hypothesis number one: character box with a symbol will have higher average gradient absolute value. It is obvious that using the distribution of the absolute gradient (see Fig. 6) we cannot build a classifier with quality over 96.5% (quality of the base method).



Figures 6. Average Gradient Absolute Values for Spaces and Symbols in Reference Dataset

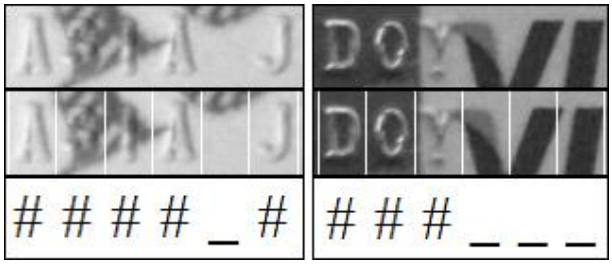
Trivial hypothesis number two: character box with a symbol will have higher dispersion of pixel intensities.
The distribution shown on Fig. 7 also does not provide the discriminative power sufficient to surpass the base method.
Besides, since these values are calculated from the same image in similar way and represent similar image properties (in slightly different ways) their errors are correlated. That means that even combination of these methods will not achieve the required quality.



Figures 7. Dispersion Values for Spaces and Symbols in Reference Dataset

A NEW APPROACH

We have decided to use a broader visual context of a symbol as an input for the new artificial neural network (2-cell NN).
A hypothesis was that the image of two conjoined characters would present more information about the background, allowing us to perform better space/symbol distinction.



Figures 8. Spaces/Symbols Detection Problem

Main feature of the suggested idea is that the NN uses both neighboring symbol images to determine the grade of similarity between the left and right parts of the input image. Hence, this NN has four output pins: two spaces, space and symbol, symbol and space, two symbols. Switching for task of similarity distinguishing between two symbols instead of direct space detection we hope to solve some complex cases (see Fig. 8).
We should notice that we did not expect to gain huge quality increase using only this method. Instead, we are creating a method with different error distribution and combining it with other techniques. The idea of the combination of different methods is not new and is well described in literature (Sergey Tulyakov et al., 2008).
It was not the first experiment for space/symbol detection using a separate method. There were attempts to train NN with two output pins, but with one character image as an input. The main difference between the base method and the 2-cell NN was the output alphabet. It could be the reason why we could not achieve quality improvement.
We have implemented proposed approach within development of a recognition system for credit cards. As a reference dataset, we have used a set of different images of 4320 cards containing about 130000 name field characters. All neural networks were trained on a separate dataset of about 230000 character images with ground truth.
We have prepared a special train dataset consisting of every pair of neighboring symbols and spaces with proper labels and used it to train our NN with four output pins.
At the recognition stage, we have used conjoined characters as an input for the NN and compared results with the base method that tries to recognize every character separately.

RESULTS

We have obtained the following results (Table 1).

Table 1. Per-Symbol Result

	Spaces	Symbols	Overall
Letter NN	93.6%	99.8%	96.5%
2-cell NN	94.3%	99.6%	96.8%
Samples Count	68246	61903	130149

This table shows that 2-cell NN has better quality in overall and a different errors distribution. It shows better space detection precision and a slight decrease in symbol detection precision. Using only this NN allows us to increase recognition quality a bit, but our goal is the combination of the methods.
What is very important here is errors distribution, which is different (Table 2).

Table 2. Errors Correlation

	Spaces	Symbols	Overall
Errors 1	4392	141	4533
Quality 2	44.7%	29.8%	44.3%
Errors 2	3893	241	4134
Quality 1	37.6%	58.9%	38.9%

First two rows of the table represent the amount of errors of the letter NN and the quality of the 2-cell NN over these samples. Third and fourth rows show the same, but for the 2-cell and letter NNs respectively.
This allows us to combine methods in order to create a new one with better quality. We take both estimations and simply make a new one as their weighted combination. This is how we use the idea in the resulting system.
In the Table 3 we have per-image overall recognition quality of the system. 2-cell version uses only 2-cell NN for space detection, whereas combined version represents a combination of letter and 2-cell networks in the task of space detection.

Table 3. Per-Card Result

	Quality, %	Quality, count
Base version	96.39	4164
2-cell version	96.46	4167
Combined version	97.06	4193

It is also worth noting that space detection problem is not the only source of recognition errors. They come from all subsystems in various amount. In our case the greatest problem is OCR itself due to the input image imperfections, but it is a topic for a different paper.
We have successfully applied this technology in existing and working system of credit card recognition.

CONCLUSION

As we have shown, information about background from a broader zone helps to recognize some of the background elements as a true space character instead of a false symbol. This idea can be improved using even wider zone of an image or performing different transformations for the database in order to train the NN to recognize spaces on given samples of the text line region. Further extending of the zone can require much wider training dataset; this problem is a topic for a different paper.

AKNOWLEGEMENTS

We would like to acknowledge the support of the Russian Foundation for Basic Research (RFBR) (grants 13-07-12171 and 13-07-12178).

REFERENCES

Anil K Jain, Øivind Due Trier, Torfinn Taxt. 1996. “Feature Extraction Methods For Character Recognition - A Survey”. *Pattern Recognition*. No 29(4); 641-662;

Bong-kee Sin , Jin H. Kim. 1997. “Ligature modeling for online cursive script recognition”. *IEEE Trans. Pattern Anal. Mach. Intell,* 623 - 633

Datong Chen, Hervé Bourlard, Jean-philippe Thiran. 2001. “Text Identification in Complex Background Using SVM”. *Computer Vision and Pattern Recognition – CVPR*. Vol. 2; 621-627;

Jagroop Kaur, Dr.Rajiv Mahajan. 2014. “A Review of Degraded Document Image Binarization”. *International Journal of Advanced Research in Computer and Communication Engineering.* Vol. 3, Issue 5, May 2014;

Frank Rosenblatt. 1957. “The Perceptron, A Perceiving and Recognizing Automaton”. *Project Para Report* No. 85-460-1, Cornell Aeronautical Laboratory (CAL).

Fukushima, Kunihiko. 1980. “Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position”. *Biological Cybernetics* 36 (4); 193–202;

Katherine L. Bouman, Golnaz Abdollahian, Mireille Boutin and Edward J. Delp 2011. "A Low Complexity Sign Detection and Text Localization Method for Mobile Applications". *IEEE TRANSACTIONS ON MULTIMEDIA*. VOL. 13, NO. 5; 922 – 934;

McCulloch, Warren, Walter Pitts. 1943. “A Logical Calculus of Ideas Immanent in Nervous Activity.” *Bulletin of Mathematical Biophysics*, No 5 (4); 115–133;

Natalya Skoryukina, Dmitry P. Nikolaev, Alexander Sheshkus, Dmitry Polevoy. 2015. “Real time rectangular document detection on mobile devices”. *Proc. SPIE* 9445, Seventh International Conference on Machine Vision (ICMV 2014), 94452A (February 12, 2015); doi:10.1117/12.2181377;

Sergey Tulyakov, Stefan Jaeger, Venu Govindaraju, David Doermann 2008. “Review of Classifier Combination Methods”. Machine Learning in Document Analysis and Recognition. Studies in Computational Intelligence Volume 90; 361 - 386;

Sholomov, D.L.; V.V. Postnikov; A.A. Marchenko; A.V. Uskov. 2005. “OCR result post-processing using partially defined syntax”. *Proceedings of Institute for Systems Analysis RAS, Intellectual information technologies - Concepts and Instruments,* Vol. 16, Moscow, KomKniga, 146-165 (In Russian).

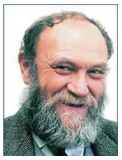
Slavin, O.A., I.M. Yanishevsky, E.B. Bogdanova. 2011. “An algorithm for recognition result validation using the dictionary”. *Information technologies and computing systems*, No. 3, 30-36 (In Russian).

Tao Wang, Wu, D.J., Coates, A., Ng, A.Y. 2012. “End-to-end text recognition with convolutional neural networks”. *IEEE*, Pattern Recognition (ICPR), 21st International Conference on; 3304 – 3308;

Vivek Shrivastava and Navdeep Sharma. 2012. “Artifical nerual network based optical charater recogniyion”. *Signal & Image Processing: An International Journal (SIPIJ)* Vol .3, No 5; 73 – 80;

AUTHOR BIOGRAPHIES

ALEXANDER SHESHKUS was born in Stary Oskol, Belgorod region, Russia. He has entered the Moscow Institute of Physics and Technology, where he was studying physics, discrete mathematics and Computer Science. He obtained his Masters degree in 2011. Since 2013, he has been software developer at the Smart Engines Ltd. His research activities are in projective invariant image segmentation. His e-mail address is: `astdcall@gmail.com`.



VLADIMIR LVOVICH ARLAZAROV was born in Moscow, Russia. Graduated from Moscow State University (MSU) in 1961, where he studied physics and mathematics. He obtained his Ph.D. degree in 1968 from MSU and Doctor of Sciences degree in 1988. He founded the company Cognitive Technologies in 1993. He became the member of Academia Europaea in 2007 year. Now Vladimir Lvovich Arlazarov is head of the Department of Innovation and High Technology at the Institute for Systems Analysis of Russian Academy of Sciences (RAS). His research activities are in the areas of Optical character recognition. His e-mail address is: `arl@isa.ru`.