On-Board Recognition System for Reading Foreign Signs in Real-Time

Ruslan Tetuev[A], Sergey Alyoshin[B]

***Abstract* — Today we witness the rapid growth of digital information and its main characteristics like image size and frame rates. This explains nowadays tendencies in the data processing toward analytics. Analog-to-digital processes become more precise and continuous, so now the data don't look like a set of discrete and extremely noisy values. In late 80's digital images were just byte codes, but now they are usually represented by smooth functions, like cosines in JPEG, MPEG. Following that tendency we propose a solution for drivers, who cannot read information signs in different languages – an on-board system equipped with high-resolution cameras for capturing these signs on the fly and a PC to process the images in real-time. But first some improvements should be made in the theory to handle practically such a huge stream of data with cheap on-board PCs, i.e. we should substitute every characters matrix for its shape invariants, what works much better for high resolution images. In this paper we discuss the basics of this idea.**

Key Words — driving, on-board system, eye tracking, road information signs, machine vision, text-to-speech

1. Introduction

Investigating the evolution process of nowadays hardware and software marketing it is easy to predict new tendencies in electronics and to reveal new trend in ongoing research activities. Today the data retrieved using analog-to-digital conversions (ADC) have higher quantization resolutions, greater sampling rate and less amount of noise. The data of the "continuous origin" look more continuous and smooth, less discrete and noisy. Obviously, these tendencies could not be ignored by scientific researchers, who had to correct the way they process the data and to propose new methods and approaches when it had to be radically changed. For example, in 80's most of digital images were simply code sequences for a finite set of colors, compressed just like text strings (GIF, PIX, TIFF), but now digital images are usually represented by smooth continuous functions, e.g. cosine functions in JPEGs, MPEGs, etc.

A long time ago, before the first personal computer was invented, in the Radio Corporation of America one of the first text-to-speech devices was developed and built [1]. This happened in early 1949; however one of the earliest applications of Optical Character Recognition (OCR) was patented even earlier in 1929. These ancient OCR devices were too expensive and frankly not so effective and handy.

However sci-fi writers of that age were sure that in the nearest future we should expect a rapid growth in this area of science and technology, they artistically described the urban roads flooded with vehicles being driven by smart robots [2]. Today is 2010, we have indeed powerful multi-core processors and high-resolution cameras with high-frequency shooting, but the problems of self-driving cars are still quite a challenge. The authors of this paper believe that the core of this like problems lies in the lags between rapid ADC devices improvements and comparatively slow theoretical enrichments. Here we propose an approach that could be used to develop an electronic assistant for drivers, designed for capturing the roads' information signs written in different languages, to convert them into text, translate the text and to read it out loud in driver's native language.

2. The technological problem

Electronic in-car devices are one of the vastly developing areas of computer science and technology. Mobile devices for Automatic Number Plate Recognition (ANPR) become probably the first application of computer vision [3] used on police patrol cars. Such a device is capable to extract the number plate's images from the incoming photos of real world and convert them to the text in real time. Police officers usually get an alarm signal when some interesting license plate retrieved from the environment.

Our main problem is to develop a devise for getting road information signs out from the environment's photos and to narrate them in driver's native language while he/she is traveling overseas by car. Unlike the case of pedestrians, who have much time to translate unknown texts and have various abilities to inquire what all these unfamiliar signs actually means, in the driver's case we cannot do the same, in that case we are restricted in time and, for the reasons of driving safety, we should have focused more on the road.

Even though this problem is new and unsolved it should be mentioned here that we already have some interesting applications for mobile devices today, that could be used by tourists for translating restaurant menus, tickets, bills, etc., like in [4]. However there are still a lot of problems to overcome before we would be able to propose an in-car devise with the same functionality. The main and the most obvious problem is, again, a lack of the

A - Dr. Ruslan Tetuev, e-mail: ruslan.tetuev@gmail.com
B – Sergey Alyoshin, e-mail: 4memph@gmail.com
The authors are with Laboratory of Data Processing, Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Instituskaya 4, Pushchino, Moscow reg., Russia Fed.

time for image processing. In the paper below we will discuss how much we can benefit from changing a focus of image analysis, what theoretical improvements should be accomplished, and why these improvements cannot be done by scholars previously for solving the same technological problems.

3. The OCR evolution: Dots vs. Shapes

Optical Character Recognition (OCR) – one of the first applications of Pattern Recognition and Image Analysis. Ancient machine vision programs operated mainly with extremely small amount of data, in early 80's most of the electronic copy machines were oriented to and restricted by human recognition abilities. For instance, default fonts of PCs consisted of monochromic characters represented by 8x8 matrices (patterns) and their printed versions were made up of 8-20 dots per symbol. For a long time neither technical, nor economical reasons permit us to pass these restrictions through and to retrieve better representations for visual copies of text information.

Needless to say that the specifics of the data influence the way we manage with it. For example, if for a scanned text copies we have about 8x8 pixels per a character, then we should generate a set of dot-patterns for any characters (alphabet patterns) and then to compare every extracted unknown dot-matrices with all patterns from the alphabet, dot-by-dot. And this is pretty time-consuming operations, the whole process is complicated and sophisticated, but yet efficient when we had to manipulate with small patterns. This like approach is very good when there is no way to estimate the shape and geometry of extracted dotty spots, which should be interpreted as characters or symbols from the considering alphabet, or simply meaningless blobs.

Nowadays the OCR technologies are getting focused on more 'geometrical' approaches, because it is much faster to extract the contours of all blobs on the image plane, to estimate geometrical characteristics of these spots and then to compare small set of these values instead of performing lengthy pixel-to-pixel matching process. Indeed, today we have high-resolution scanners/cameras, it is not a problem anymore to get images of good quality, containing highly detailed characters, even when it retrieved for small fonts. So, if it is easy now to extract the contours of characters, why not to change the way we are processing the images radically for the sake of recognition speeding. This is what they call shape-based recognition approach, broadly used previously in handwriting recognition, reCaptching, etc. This approach allows us to obtain some promising results and it is much more suitable for the problem we are trying to solve, as demonstrated below.

4. Basics of shape-based OCR

Any contour of the object detected on the image plane we can easily represent as a planar curve: $[x(t), y(t)]$, and these functions in turn could be approximately represented by a set of $2N$ spectral coefficients $[A_n^x, A_n^y]$:

$$\begin{cases} x(t) \approx \sum_{n=0}^{N} A_n^x T_n(t)\,; \\ y(t) \approx \sum_{n=0}^{N} A_n^y T_n(t)\,, \end{cases}$$

where $\{T_n(t)\}$ – is a set of orthogonal functions [6]. Spectral approximations of a contour could be much more certain, correct and detailed when the N value is high, and vice versa for less amounts of spectral coefficients we get very rough models of the original shape (see Fig. 2).

Fig. 1: The STOP sign in the Russian city St. Petersburg.

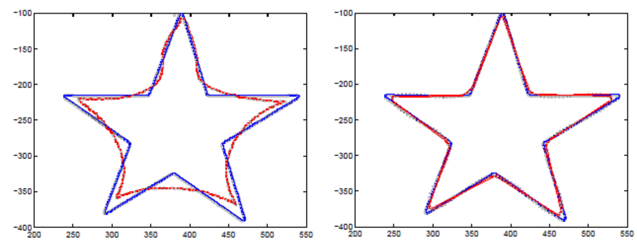

Fig. 2: Spectral approximations (red) of the same contour (blue) with 15 (left) and 25 (right) coefficients.

After some normalization of the shapes is done we can expect similar coefficients' sets for similar contours, which could be estimated for instance through the degree of correlation between these spectral sets. The best part of the shape-based approach – we can perform all these steps in a very rapid manner, including the final step when we should search through the whole databank of spectral sets, obtained in advance for all known characters and symbols.

5. Performance Experiments

Traveling in different countries the authors have gathered up some interesting collection of traffic information signs written in different languages. Since many of these shots were taken by the same camera we decided to use them for performance experiments. That was a Sony Ericsson Z550i camera phone with only a 1.3 Mega pixels camera. In spite of that the shots were taken from vehicles the quality was enough to detect, extract and read not only the information signs, but also the advertisement texts. As you can see on the Fig. 3 there is a weight limit sign, which was perfectly read by our OCR program (desktop version), in addition this program extracted and read the ads written in German: "Euro Mietwagen," what means "Euro Rental Cars." In accordance with our performance tests it will take about 5-7 seconds for a modern mobile device's applications to get the same results.

Fig. 3: The weight limit road sign in the German city Baden-Baden.

6. Future Work

In the near future the authors are planning to finish their Java-based application with recognition functionalities designed for mobile devices with an inbuilt cameras using Java2ME technologies. Nevertheless the obtained results are very promising there are many additional capabilities to accelerate the algorithm using high speed low level Image Processing as it shown for example in [7].

References

[1] "*Reading Machine Speaks Out Loud*," February 1949, Popular Science

[2] "The *Car That Drives Itself*," May 1958, Popular Science.

[3] J. Setchell, "*Application of Computer Vision to Road-Traffic Monitoring*," PhD Thesis, University of Bristol, 1997.

[4] I. Haritaoglu, "*Scene text extraction and translation for handheld devices*." In Proc. IEEE CVPR 2001, volume 2, pages II–408–II–413 vol.2, 2001.

[5] S. Kim, S. Oh, M. Kang "*Mathematical image processing for automatic number plate recognition system,*" J. KSIAM Vol.14, No.1, 57–66, 2010.

[6] S. Alyoshin, F. Dedus, R. Tetuev, "*Spectral approach to the estimation of affine invariants,*" Mathematical Methods for Pattern Recognition 14, 2009.

[7] L. Xu, Zh. Wang, H. Wang, A. Shi, Ch. Li, "*A J2ME-Based Wireless Intelligent Video Surveillance System Using Moving Object Recognition Technology,*" cisp, vol. 2, pp.281-285, 2008 Congress on Image and Signal Processing, Vol. 2, 2008.