Real-time Vehicle Number Plate Recognition from Video Sequences

PAOLO CASTELLO, ENRICO DEL NINNO, ENNIO OTTAVIANI
Research and Development Department
Elsag S.p.A.
Via Puccini 2, 16154 Genova
ITALY

Abstract: The problem of recognizing the number plate from video image sequences of moving vehicles is addressed. We found that an efficient approach is to process each image of the sequence in order to extract character information, either complete or partial, and to merge them in a temporal post-processing step. This ensures that local errors in segmentation or character recognition may be recovered during the permanence of the vehicle inside the camera field of view. Moreover, there is no need of an external device to trigger image acquisition, because the entire image sequence is processed. This allows easy integration in existing traffic control system.
The general approach may be organized as follows. The first stage analyzes the image difference between consecutive frames in order to verify the presence of a moving vehicle. Then each frame showing motion is processed in order to locate plate candidates, defined as sub-images likely to contain text. The sub-images are segmented in order to find character-like objects. Such objects are passed through an OCR stage and the resulting ones are analyzed with respect to the allowed spatial distribution and syntax of national number plates. Here is possible to recognize the number plate either completely or partially. In both cases, alpha-numeric and spatial attributes of the selected characters are fed into a post-processing module which performs the final recognition.
IMACS/IEEE CSCC'99 Proceedings, Pages:4481-4484

Keywords: OCR, segmentation, video processing, plate recognition, real-time, context analysis

1 Introduction

Vision-based vehicle number plate recognition systems have today a lot of applications in traffic surveying and monitoring, e.g. tracking of stolen cars, controlling access to car parks and limited traffic zones, gathering traffic flow statistics [1]. In co-operation with other kind of technologies, microwaves for instances, they can be used also for reinforcement of automatic toll collection systems. Most existing systems require an external device in order to trigger image acquisition from a CCD camera, and often use special hardware to implement computational intensive image processing algorithms. The approach proposed here deals with the processing of the entire video sequence and so does not require any external device for triggering. A fast algorithm is used in order to quickly extract the image regions which are more likely to contain the plate. This allows consistent reduction of image processing time, usually lower than 40 msec on a PC platform, and the ability to process a video sequence in real time without any delay. Moreover, because of the use of many images to recognize the number plate of a single car, high recognition rate may be achieved by temporal integration of partial data.

2 System architecture

The system architecture of the number plate recognition system may be sketched as follows:

- A pre-processing block performing a screening of the video sequence in order to selects images or sub-images showing motion;
- An image processing engine performing plate location, character segmentation, OCR and context verification;
- A temporal post-processing block performing the fusion among character data coming from different images of the sequence.

2.1 Sequence pre-processing

The pre-processing of the sequence consists in the estimation of the background image, which is recursively updated as soon as a new image is available. The difference images is computed and thresholded in order to identify regions showing large image variation in gray levels. If the total area of the detected region is null or lower than a fixed threshold, the image is discarded, otherwise the coordinates of the bounding box containing all regions are evaluated in order to drive the following stage of the process. Being a very computational expensive task, background estimation and image difference evaluation are carried out at a lower resolution with respect to other low level image processing functions. Dealing with typical *768x256* images, a downsizing of *16* allows good screening performances without affecting sensibly the total processing time.

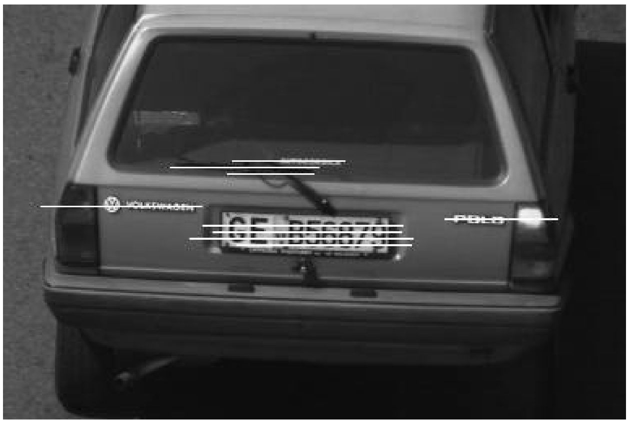


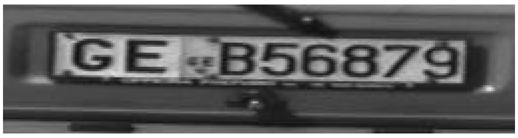
Fig.1 - Typical car image with superposed line segments

2.2 Image processing engine

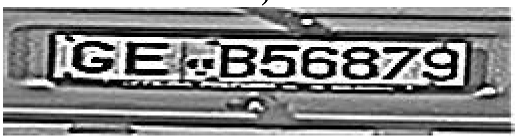
The image processing engine is the core of the system, because it is responsible to extract character information from the raw images. At this level there are no conceptual differences between it and the large number of plate recognition algorithms operating on single images already available. However, in order to achieve maximum speed and to allow real-time sequence processing, we have to design an algorithm solution which is a good compromise between accuracy and computational complexity.

2.2.1 Plate location

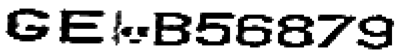
In order to reduce the computational complexity, the best approach is trying to locate the plate as soon as possible, in order to avoid low-level processing of most part of the image. The approach implemented here is a version of the "plate signature" algorithm proposed in [2]. The underlying idea is based on the fact that the lines of the plate have a clear structure in terms of gray-level distribution, if observed along an horizontal segment passing through them. There must be a sequence of minima and maxima with good contrast and almost regularly spaced. These characteristics (number of points, relative distances, gray level differences) may be evaluated and compared with a statistical reference in order to decide if a given line is passing through a number plate or not. If so, the most probable end points of the cutting segment can be also estimated.
So our approach consists in cutting the image (or the sub-images) by a set of equally spaced horizontal lines and evaluating a set of segments which exhibit such kind of property. The spacing between adjacent cutting lines has to be adjusted with respect to expected character height. Segments are then clustered in order to define the bounding box of the candidate number plates.



a)



b)



c)

Fig.2 - Processing stages: a) plate location, b) convolution, c) character segmentation

2.2.2 Character segmentation

In order to keep in account the extremely variable lighting conditions of number plated in an outdoor environment, we adopt a segmentation algorithm based on image convolution with a *2D* mask approximating the laplacian of gaussian [3]. This is a well known method to extract dark objects on light background (or vice-versa). The size of the gaussian is chosen according to the expected character thickness. In common applications we adopt *7x7* convolution masks with a positive *3x3* kernel, operating only on candidate number plates generated by the previous location task. A dyadic mask is used to avoid products (only bit shifts are required).
The convolved image is then adaptively thresholded in the following way. The *RMS* laplacian image value L_0 is estimated, then the following rules apply:

- if $L(x,y) < -L_0 \Rightarrow$ the pixel belongs to a black/white object;
- if $L(x,y) > L_0 \Rightarrow$ the pixel belongs to a white/black object;
- otherwise the pixel is meaningless.

The object of both positive or negative sign are then generated by a region growing algorithm, grouping

together all 8-connected pixels sharing the same sign. They are the candidate characters to be analyzed by the following OCR task.
In order to reduce the number of candidates to be recognized, it is possible to use many contextual information concerning minimum and maximum size of the characters, aspect ratio, gray level differences between interior and exterior regions of a character. Sometimes is also possible to discard one of the two sign, usually the white on black one, if the allowed number plates in a given country may be composed only by dark characters.

2.2.3 OCR

The OCR approach used here belongs to the standard statistical pattern recognition scheme. A vector of feature is evaluated for each candidate character and the distances between this vector and a set of prototypes generated during the training phase is computed.
Dealing with gray level characters, an equalization step is needed in order to normalize lighting conditions. We adopt a linear equalization procedure where only two parameters (gain α and offset β) are need to compute the new gray levels.

$$z' = \alpha \cdot z + \beta \qquad (1)$$

Gain and offset are evaluated requiring the new gray level distribution has fixed mean value and variance. Moreover, white on black characters are reversed in order to be compared with the prototypes stored in the black on white format.
The feature vector which gives the best results in term of accuracy is composed by the averaged gray level computed on a rectangular mesh adaptively spaced over the sub-image of the character. The typical mesh size is *7x7*, because it gives good recognition percentages at an acceptable computational cost.
Best recognition results have been obtained by the use of the Mahalanobis distance.

$$d\left(X, M_i\right) = (X - M_i)^T \Sigma^{-1} (X - M_i) \qquad (2)$$

where X is the feature vector. This requires the computation of a mean vector M_i and a covariance matrix Σ_i for each alphabetic and numeric class.
Distances from alphabetic and numeric prototypes are then ranked separately and the first two of each type are stored for the contextual verification.
However, if the first alphabetic (numeric) distance is too high, the character is assumed to be numeric (alphabetic). Even the second hypothesis of each type is discarded if it is much higher than the first one. Furthermore, if both first distances are too high, the candidate character is completely discarded.
The prototypes are evaluated over a data-base of gray level character images (about *10.000* instances of the *36* alphanumeric classes). Testing result shows a great accuracy in term of recognition of true character (more then *99.9%).* Dealing with false characters, the Mahalanobis distance provides generally values much higher than on true ones, so it is quite simple to discard them during the contextual verification process.

2.2.4 Context verification

The context verification process exploit both spatial and syntactic information in order to select the best hypothesis for the number plate. Let us suppose that a given country allows number plate with K characters in fixed position $P_i=(X_i,Y_i)$ (expressed in a given system of coordinates). If the image under processing contains N validated characters, the general idea is to extract all choices of K elements from N and to evaluate them both spatially and syntactically.
Spatial evaluation means the computation of the RMS fitting errors between the observed character coordinates $p_i=(x_i,y_i)$ and the transformed version of the allowed positions P_i. Affine transformations are used in order to model the deformations induced by the viewing geometry (e.g. by the distance from the camera and the viewing angles).

$$\begin{cases} x_i = a + b \cdot X_i + c \cdot Y_i \\ y_i = d + e \cdot X_i + f \cdot Y_i \end{cases} \qquad (3)$$

The parameters of the transformation are evaluated by *LMS* fitting and the residual fitting square error is assumed to be an estimate of the spatial plausibility of a given hypothesis (spatial cost).
Syntactic evaluation means the verification of the existence of the allowed alphabetic and numeric distances in each position of the number plate. The sum of such distances is taken as an estimate of the syntactic plausibility of a given hypothesis. (syntactic cost).
In order to avoid the rapid grow of possible hypothesis when $N >> K$, an early pruning approach has been used. We firstly sort the N characters in the most natural "reading order", from top to bottom and from left to right. So the possible choices of K characters may be generated and evaluated sequentially in a left-to-right Markov model [4]. Here it is also possible to impose constraints about plate size, character spacing and to allow the presence of one or more spurious characters among the true ones.
Finally all complete hypothesis are ranked according to their cost and the best one (or two if there is still

some ambiguity) is retained for temporal post-processing.
Such approach allows also the recovery of missed characters there is not a choice of *K* characters from the original *N* with a sufficiently low cost. It is sufficient to verify if it is possible to realize a good cost by putting together *K-1* characters with small fitting error an correct syntax and a new one obtained by cropping a sub-image right in the coordinates computed by the affine transformation of the missed position.
This recovery scheme allows the detection of missed characters and so the recognition of the number plate in many cases of poor image quality.

2.3 Temporal post-processing

The final temporal post-processing stage aims to extract a single number plate for a given vehicle. The presence of a vehicle is detected by the observation of a group of video frames affected by image motion and showing a sufficient number of recognized characters. Such definition allows the detection of a vehicle even if the contextual process does not extract any number plate hypothesis.
All the number plate hypothesis, with both spatial and syntactic information, are gathered from the image belonging to such group and a clustering approach is used in order to evaluate the best choice.
During the clustering, two number plate hypothesis can be merged if they have small syntactic distance (the Levenshtein distance is used to compute the distance between alphanumeric strings) and they are spatially coherent with the assumed vehicle trajectory in the image plane.

3 Experimental results

The system has been extensively tested on a tolling station of an Italian motorway. Results show a recognition of about *95%* under good lighting conditions, with an accuracy greater than *99%*. Cars are passing through the field of view of the camera at a moderate speed (about *40 Km/h*), so about *10* video frames are available for each vehicle. That means even if speed should be higher, there should be a sufficient number of frames in order to ensure good recognition rates.
Problems arise with strong shadows occluding part of the imaged scene. In such case the automatic shutter of the camera may fail to select the proper exposure time in order to get a recognizable image of the number plate.

4 Conclusions

We present an approach to automatic number plate recognition using computer vision. This allows the processing of a video sequence in real-time, so it does not need any external device in order to trigger image acquisition. The processing of more images of the same vehicle ensures a high recognition percentage even under bad lighting conditions.

References

[1] A.Ali et al., Vision based road traffic data collection, *Proc. ISATA 26th Int. Conf.* Aachen, Germany, 1993

[2] J.Barroso et al., Real-time number plate reading, *Master Thesis, Aveiro Univesity, Portugal*, 1995.

[3] R.Jain, and R.Kasturi, *Machine Vision*, Mc-Graw Hill, New York, 1995

[4] C.Nieuwoudt, R.Van Heerden, A connectionist approach to vehicle licence plate recognition, *Proc. PRASA 97, IAPR 8th South African Work. on Patt. Rec.,* pp.75-79, 1997.