Detecting, Tracking and Recognizing License Plates

Michael Donoser, Clemens Arth, Horst Bischof

Institute for Computer Graphics and Vision
Graz University of Technology
{donoser,arth,bischof}@icg.tu-graz.ac.at

Abstract. This paper introduces a novel real-time framework which enables detection, tracking and recognition of license plates from video sequences. An efficient algorithm based on analysis of Maximally Stable Extremal Region (MSER) detection results allows localization of international license plates in single images without the need of any learning scheme. After a one-time detection of a plate it is robustly tracked through the sequence by applying a modified version of the MSER tracking framework which provides accurate localization results and additionally segmentations of the individual characters. Therefore, tracking and character segmentation is handled simultaneously. Finally, support vector machines are used to recognize the characters on the plate. An experimental evaluation shows the high accuracy and efficiency of the detection and tracking algorithm. Furthermore, promising results on a challenging data set are presented and the significant improvement of the recognition rate due to the robust tracking scheme is proved.

1 Introduction

There is a need for intelligent traffic management systems in order to cope with the constantly increasing traffic on todays roads. Video based traffic surveillance is one of the key parts of such installations. Beside detection and tracking of vehicles, identification by license plate recognition is important for a variety of applications like access-control, security or traffic monitoring.

In general, license plate recognition systems consist of two separate parts. First, plates are detected within a single frame of a traffic video sequence and second, character recognition is applied to identify the characters on the plate.

Different methods have been proposed for the detection of license plates. Shapiro et al. [1] use a mixture of edge detection and vertical pixel projection for their detection module. In the work of Jia et al. [2] color images were segmented by the Mean Shift algorithm into candidate regions and subsequently classified as plate or not. The AdaBoost algorithm was used by Dlagnekov and Belongie [3] for license plate detection on rather low resolution video frames. Matas and Zimmermann [4] proposed a different approach for the localization of license plates. Instead of using properties of the plate directly, the algorithm tries to find all character-like regions in the image by analyzing an interest region detection result.

Also several approaches for recognizing the characters on the plate after successful detection were proposed. Shapiro et al. [1] use adaptive iterative thresholding and analysis of connected components for segmentation. The classification task is then performed with two sets of templates. Rahman et al. [5] used horizontal and vertical intensity projection for segmentation and template matching for classification. Dlagnekov and Belongie [3] use the normalized cross correlation for classification by analyzing the whole plate, hence skipping segmentation.

Although much scientific work has focused on recognizing license plates from traffic video sequences, surprisingly little work has been done on integrating a tracking scheme to gather additional representations of the plate for improving the recognition rate. Furthermore, in the few systems that apply a tracking scheme only simple and unstable approaches are used, as e. g. by Dlagnekov and Belongie [3] who perform tracking by simply repeating detection for building correspondences.

The main contribution of this paper is a novel framework which unifies detection, tracking and recognition of license plates in a robust and efficient way. The underlying idea is to base detection, tracking and character segmentation on the same principles which allows to provide segmentations of the individual characters for recognition in addition to accurate and robust license plate localization results in subsequent frames. The framework is presented in detail in Section 2 and an experimental evaluation is shown in Section 3.

2 License Plate Recognition Framework

This section describes the entire framework for detection, tracking and recognition of license plates from traffic video sequences. The introduced system detects newly appearing license plates from the sequence by a novel algorithm which is based on the analysis of Maximally Stable Extremal Region (MSER) [6] detection results. The concept, introduced in Section 2.1, does not require any learning scheme and is capable of detecting different types of international plates. After detection of a plate it is robustly tracked through the sequence by a modified version of the MSER tracking framework [7] as shown in Section 2.2. Therefore, for each appearing car in the video sequence a set of license plate representations is collected which is used to improve the subsequent character recognition. Finally, Section 2.3 describes how support vector machines are used to recognize the characters on the different representations of each plate and how the results are merged by a voting scheme to achieve the final recognition result.

2.1 License plate detection

The first step of every license plate recognition system is the detection of the license plates within a single frame of the traffic video sequence. We propose a novel detection scheme which is motivated by the work of Matas et al. [6]

who presented a method to learn category-specific extremal regions, i. e. the characters on a license plate, to perform detection. In contrast to their approach our scheme does not require any learning scheme and is able to detect different types of international license plates without adaption.

Our detection algorithm is based on analyzing the results of a Maximally Stable Extremal Region (MSER) detection [8]. MSERs denote a set of distinguished regions and have proven to be one of the best interest point detectors in computer vision [9]. All of these regions are defined by an extremal property of the intensity function in the region and on its outer boundary. Special properties form their superior performance as stable local detector. The set of MSERs is closed under continuous geometric transformations and is invariant to affine intensity changes. Furthermore, MSERs are detected in every scale. We predominantly exploit these properties for segmentation purposes.

In general, two variants of MSER detection can be distinguished denoted as MSER+ and MSER−. While MSER+ detects bright regions with darker boundary, MSER− finds dark regions with brighter boundary. Figure 1(a) shows an image from a traffic video sequence and Figure 1(b) and Figure 1(c) illustrate the corresponding MSER detection results as binary images. As can be seen, the license plate itself is identified as MSER+, whereas the characters on the plate are detected as MSER−.

(a) Input frame (b) MSER+ (c) MSER−

Fig. 1. MSER detection results can be used for detecting license plates in video sequences. MSER+ finds the license plate, whereas MSER− identifies the individual characters on it.

The underlying idea of our novel license plate detection scheme is to analyze the MSER+ and MSER− detection results. We are looking for a larger MSER+ region (license plate) that contains a set of smaller MSER− regions (characters). Such a combination is considered as license plate detection result. Furthermore, we verify the detection by checking if the MSER− regions are approximately equal sized, if their center points approximately lie on a line and if the height of the MSER+ is in the range of the average MSER− height. After verification, the MSER+ is returned as license plate localization result and additionally segmentations of the characters are provided by the corresponding MSER− detections.

Although the detection process is simple, it is effective and allows stable and accurate detection of license plate in complex scenes. An exemplary result is shown in Figure 2(a). In this example 226 MSER− and 688 MSER+ regions are found, but only one set fulfills the proposed criterions.

For comparison, Figure 2(b) shows the result of an AdaBoost detector based on Haar-like features [10]. As can be seen the boosting framework returns wrong and multiple detections which need to be significantly post-processed, as e. g. by using non maximum suppression to remove multiple detections. Our result is also more accurate as the bounding box provided by the boosting variant. Furthermore, training of an AdaBoost based classifier is a rather complex procedure, and in this case was especially trained on Austrian license plates. Our approach does not require any learning scheme and is able to identify different types of international plates, because the simple criterion is fulfilled for almost all of them. Figure 3 shows detection results for different international types.

(a) Proposed algorithm   (b) AdaBoost result

Fig. 2. License plate detection results. Figure (a) shows the single detection of the proposed algorithm indicated by the white border. 914 MSERs are detected, but only one set of regions fulfills the described criterion. Figure (b) shows the wrong and multiple detections of an AdaBoost algorithm.

2.2 Tracking of license plates

After detection of a newly appearing car in the traffic video sequence by its license plate, a robust tracker is applied to increase the number of character representations for the subsequent recognition step. In general, any tracker can be used but we propose to apply a modified version of the MSER tracking framework introduced by Donoser and Bischof [7] which has some advantages in contrast to other tracking schemes. First, it is efficient and stable and can be adapted to our specific requirements. Second, it provides an accurate segmentation of the license plate and third, in addition to the tracked plate it also returns

(a) Input image (b) Characters (MSER-) (c) Plates (MSER+)

Fig. 3. Detection results on international license plates where (b) shows the segmented signs detected as MSER- and (c) the plate detected as MSER+. As can be seen the detection algorithm can be applied to any international type, because the criterion used in the detection algorithm is valid for all cases.

segmentations of the individual characters on the plate (MSERs), thus tracking and segmentation are handled simultaneously.

The MSER tracking framework was designed to improve the stability and speed of MSER detection results in video sequences. The tracker has to be initialized by passing the region to be tracked R_t detected in image I_t of the sequence to the framework. The first step of the algorithm is to propagate the center point of the region R_t to the next image I_{t+1} and to crop a region-of-interest (ROI) around it from the image. Then a data-structure named component tree [11] is built for this ROI. Every node of the component tree contains one candidate region C^i_{t+1} for the tracking and the algorithms looks for the node which is most similar to the region R_t. The best fit is identified by comparing feature vectors that are built for each of the nodes of the component tree C^i_{t+1} and the input region R_t. The candidate C^i_{t+1} with the lowest Euclidean distance between its feature vector and the one of the region R_t is taken as tracked representation. The features calculated are mean gray value, region size, center of mass, width and height of the bounding box and stability. All of these features are computed incrementally [12] during creation of the component tree. Thus, no additional computation time is required. After detection of the new representation the described steps can be repeated for tracking the region through the entire sequence.

The original MSER tracking framework was designed for tracking arbitrarily shaped region, but we adapt the method to our specific requirements of tracking license plates. In our framework, the MSER tracking algorithm is initialized by the result of the license plate detection algorithm presented in Section 2.1. Because we focus on license plates we reformulate the feature comparison approach by replacing the Euclidean distance computation by a simpler, but more effective equation based on comparison of two distinct features, the size and the rectangularity of the region. Thus, in our framework the tracked representation is found by calculating a distance value $\Delta(R_t, C^i_{t+1})$ for every candidate region C^i_{t+1} by

$$\Delta(R_t, C^i_{t+1}) = \frac{abs\left(|R_t| - \left|C^i_{t+1}\right|\right)}{|R_t|} + \left(1 - \vartheta\left(C^i_{t+1}\right)\right), \tag{1}$$

where $|\ldots|$ denotes the size of the region and $\vartheta(\ldots)$ is the rectangularity. Then, the candidate C^i_{t+1} with the lowest $\Delta(R_t, C^i_{t+1})$ value is taken as final representation. Tracking is considered as valid as long the minimum distance value Δ is below a fixed threshold. By repeating the presented steps the license plate is tracked through the sequence and the detected MSER- regions within the tracked plate are provided as segmentations of the individual characters. Furthermore, the tracking scheme is also used for discarding false positives of the detection step by removing non-moving or unstable plate tracks.

Figure 4 shows two frames of a traffic video sequence, where accurate license plate detections are provided in addition to the segmentation of the eight characters on the plate.

(a) Frame 1 (b) Frame 15

Fig. 4. Illustration of license plate tracking. The images show a traffic scene and the segmented characters on the license plate are highlighted in white.

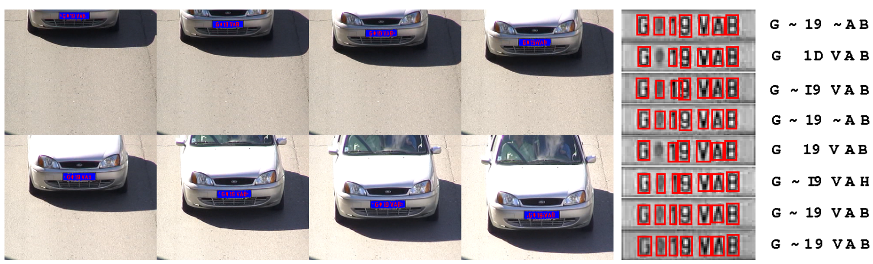
2.3 Character recognition

The final step of our framework is to recognize the individual characters on the detected plates based on support vector machines (SVMs). SVMs were first described by Vapnik [13] and have proven to be an efficient tool for classification tasks as in optical character recognition (OCR) of handwritten digits or license plate contents [14]. For an introduction to SVMs and other kernel methods see for example [15].

In general, SVMs are designed for binary classification problems. Because character recognition is a multi-class problem we apply a method based on a combination of several binary SVMs. The strategy is called *one-vs-one* where for each pair of output classes an individual SVM is trained, resulting in a total number of $n \cdot (n-1)/2$ classifiers. Then all classifiers are evaluated, the votes are summed up and the class with the maximum number is chosen.

First the provided character segmentations are aligned, and then the *one-vs-one* approach is used to classify each character independently of all the others. The presented tracking approach provides several license plate representations for every car and therefore, we also have several classification results for every character on the plate. A majority voting scheme is then used to determine the final character recognition result for every car.

Figure 5 shows a sequence of tracked license plates, the segmented characters and the corresponding classification results. As can be seen, the single image based recognition provides wrong assignments, but the subsequent majority voting scheme ensures that the final character classification is correct. Therefore, the power of our framework clearly comes from the stable tracking approach which allows to combine a sequence of single image based recognition results into the final classification.



(a) Tracking Sequence (b) Classification Results

Fig. 5. Recognition is improved by using several license plate representations provided by the tracking scheme. The final result based on a majority voting for this plate sequence is `G~19VAB` which matches with the real plate number.

2.4 Framework

Our framework is able to analyze traffic video sequences in real-time. It detects newly appearing cars by localization of their license plate. After a one-time detection, the plate is robustly tracked and several representations of the license plate are collected. Until tracking fails the available repeated segmentations of each character on the plate are used to improve the recognition rate along the complete tracking sequence and to determine a final result by the majority voting scheme. The running times of the individual steps of the concept for analysis of a video sequence of size 352×288 are shown in Table 1.

	Detection	Tracking	Recognition
Running time	$70ms$	$5ms$	$6ms$

Table 1. Running time per image of the individual steps of the framework for analyzing a video stream of size 352×288.

3 Experimental Evaluation

We evaluated our framework on a challenging traffic video sequence in the type of Figure 5(a) which was filmed from a footbridge. The provided resolution was 352×288 and therefore the characters on the plate only have an average size of 9×6 pixels in the sequence.

Section 3.1 describes how the necessary training data for character recognition was acquired. Finally, Section 3.2 demonstrates that promising results are achieved on the challenging data set and shows how the recognition rate is significantly improved by using several plate representations provided by the tracking scheme.

3.1 Character image database

The support vector machines are trained on approximately 2700 manually labeled images of characters which were automatically extracted from high resolution images of parked cars using the license plate detection algorithm. According to the resolution of the test video sequence we resized all character images to 9×6 pixels.

3.2 Recognition results

To evaluate the quality of the proposed framework we used it for recognizing the license plates of 109 cars passing the video sequence area. Due to the low resolution and changing lighting conditions the average recognition rate for independent classification of every character in every detected license plate is only 80.74. As a consequence, in a single image based license plate recognition approach on average only 23.23% of the cars are totally correct classified. But this rate is significantly improved by analyzing additional plate representations provided by the tracking scheme and combining the corresponding recognition results by the presented majority voting scheme. Figure 6 analyzes the increase of the recognition rate by using more representations. As the number of tracked plate representations gets close to 70, the percentage of totally correct classifications reaches more than 90%.

By using all available representations for recognizing the plate of every car (4880 plates) our framework classifies 94.65% of all cars totally correct which is a promising result for such a challenging data set. Please note, that no postprocessing, like checking the validity of license plates according to syntax restrictions,

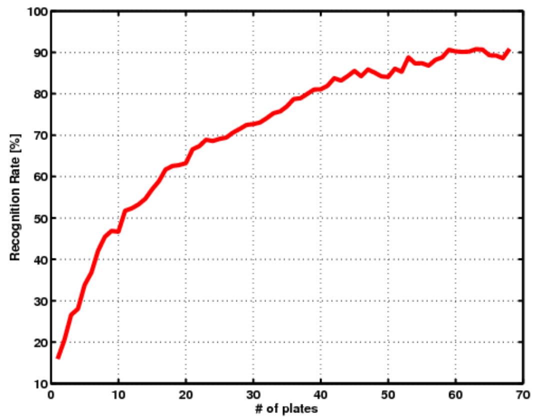


Fig. 6. Illustration of improvement in recognition rate achieved by analyzing several license plate representations provided by the tracking scheme.

is used. Table 2 analyzes the recognition rate on plate level, i. e. the number of totally correct classified plates, and character level, i. e. the number of correctly classified characters, for the single image based approach in comparison to our tracking variant. As can be seen analyzing several representations instead of a single one significantly improves the recognition rate.

	Single image approach	**Tracking approach**
Character level	80.74%	**97.16%**
Plate level	23.23%	**94.65%**

Table 2. Recognition rates of totally correct classified plates (plate level) and correctly classified characters (character level) for a single image based approach in comparison to the tracking based classification.

4 Conclusion

This paper introduced a novel framework which allows detection, tracking and recognition of license plates. Detection is handled by analysis of Maximally Stable Extremal Region (MSER) detection results and does not require any learning scheme. We introduced a robust tracking scheme, which provides accurate license plate localizations and segmented characters simultaneously. The experimental evaluation showed that promising results are achieved on a challenging data set and that the robust tracking approach significantly improves the recognition rate. Furthermore, due to the high efficiency of the individual components, the framework can be used for real-time traffic video sequence analysis. To make the proposed framework applicable in industrial scenarios, we also ported it to an DSP-based embedded platform. Although the experiments were all performed on a desktop computer, the results also hold for a fully embedded implementation.

References

1. Shapiro, V., Gluhchev, G., Dimov, D.: Towards a multinational car license plate recognition system. **17**(3) (2006) 173–183
2. Jia, W., Zhang, H., He, X., Piccardi, M.: Mean shift for accurate license plate localization. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC). (2005) 566–571
3. Dlagnekov, L., Belongie, S.: Recognizing cars. Technical Report CS2005-0833, UCSD University of California, San Diego (2005)
4. Matas, J., Zimmermann, K.: Unconstrained licence plate and text localization and recognition. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC), Vienna, Austria (2005) 572–577
5. Rahman, C., Badawy, W., Radmanesh, A.: A real time vehicle's license plate recognition system. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS). (2003) 163–166
6. Matas, J., Zimmermann, K.: Unconstrained licence plate detection. In: Proceedings of International Conference on Intelligent Transportation Systems (ITSC). (2005) 572–577
7. Donoser, M., Bischof, H.: Efficient maximally stable extremal region (MSER) tracking. In: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR). (2006) 553–560
8. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proceedings of British Machine Vision Conference (BMVC). (2002) 384–393
9. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) **27**(10) (2005) 1615–1630
10. Zhang, H., Jia, W., He, X., Wu, Q.: Learning-based license plate detection using global and local features. In: Proceedings of International Conference on Pattern Recognition (ICPR). (2006) 1102–1105
11. Najman, L., Couprie, M.: Quasi-linear algorithm for the component tree. In: SPIE Vision Geometry XII. Volume 5300. (2004) 98–107
12. Matas, J., Zimmermann, K.: A new class of learnable detectors for categorisation. In: Proceedings of Scandinavian Conference of Image Analysis (SCIA). (2005) 541–550
13. Vapnik, V.N.: The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA (1995)
14. Zheng, L., He, X.: Number plate recognition based on support vector machines. In: Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06), Washington, DC, USA, IEEE Computer Society (2006) 13
15. Schölkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge, MA, USA (2001)