FEATURE BASED RECOGNITION OF TRAFFIC VIDEO STREAMS FOR ONLINE ROUTE TRACING

Christoph Busch†, Ralf Dörner†, Christian Freytag‡, Heike Ziegler‡

†Fraunhofer Institute for Computer Graphics, ‡Computer Graphics Center
Darmstadt, Germany
Email: {busch, doerner, freytag, heike}@igd.fhg.de

Abstract — In this paper an approach is presented for the acquisition of route related data by means of recognizing vehicles that pass different survey points. This recognition approach essentially relies on vehicle license plate matrices that are provided by standard video cameras. A recognition pipeline is conceived where the main steps are a frame selection method, a segmentation technique for the localization of plate matrices, a feature extraction method and a recognition process. The single steps are discussed in detail and recognition rates that have been achieved in a field trial are presented.
Keywords: Computer Vision, motion image analysis, feature extraction, object recognition, corresponding problem, traffic analysis, route tracing

I. INTRODUCTION

Due to upcoming low cost hardware and the progress in algorithmic research, Computer Vision has become a promising base technology for traffic sensoring systems. Various research projects [1,5] pointed out advantages of video based technology over conventional systems. Among these advantages are portability, ease of installation, reliability, long lifetime and low cost. These advantages are standard features in integrated video based traffic analysis systems [3,6].

Another major new feature is the acquisition of route related data [4], which will be elaborated within this paper. The term route related data describes the combined processing of data stemming from different survey points. Linking the points of recognition to continuous routes, the traffic flow statistics can be computed from the data acquired. This data represents the current traffic situation and can be optimally used for the control of variable direction signs.

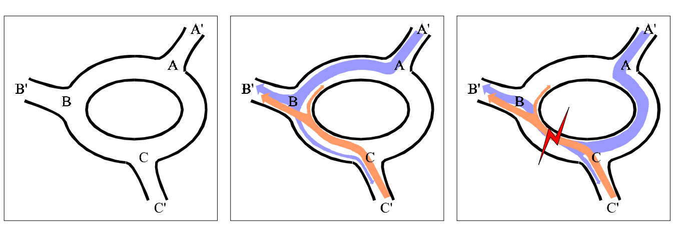

Fig. 1: Traffic situation around a major city. The modified route direction towards line A-C produces a traffic jam on line C-B

The importance of the traffic flow survey for an optimal trip distribution is illustrated in Figure 1. This figure represents the traffic situation around many major cities. Assume dense traffic running on the long-distance route from A'-B-C' (Figure 1a). As soon as traffic increases, local detectors on route segment A-B (in Figure 1b) will register higher numbers of vehicles per time. Dynamic traffic control systems would modify variable road signs in a way that the assumed main traffic flow would then run on the route segment A-C. This control reaction is contra productive if most of the vehicles are running from A' towards B' (see Figure 1c), since superposition of the diverted traffic with the dense traffic flow on line segment C-B will then lead to a complete traffic jam.

The major subject of this paper will be an approach for the acquisition of route related data by means of recognizing vehicles that pass different survey points. This recognition approach essentially relies on vehicle license plate matrices that are provided by standard video cameras.

II. IMAGE PROCESSING PIPELINE

As introduced in Section I, the survey of net-linked traffic data relies on recognition of single vehicles at different points. The recognition is accomplished by using features that characterize a vehicle. It can be shown that features extracted from wide-angle video frames (showing the entire highway) such as vehicle class (van, car, motor bike), color and predicted arrival time are not sufficient for reliable vehicle recognition. The analysis of license plates as a feature contributor is therefore considered in our approach. In this context, video recordings which show only single lanes are used. This section gives an overview of how the video recordings are processed in an image processing pipeline.

An ordinary video stream of theoretically infinite length is used as input data for the pipeline at each survey point. A number of frames are redundant since they do not contain any relevant information (e.g. showing the pavement and no vehicle). Therefore, the task of the first step, the relevant frame selection, is to filter frames of a video sequence that include a vehicle for further processing.

The second step, the segmentation process, is performed in two substeps. In the first substep, the image area roughly containing the license plate is cut from the entire frame. Secondly, a finer segmentation is executed in order to obtain the character field. Due to our application constraints we have developed a knowledge based solution for license plate segmentation that relies on the detection of horizontal and vertical plate edges. The segmentation of plate images is influenced by plate properties and local illumination conditions (e.g. shadows). In order to provide comparable information vectors for further processing, an additional segmentation and normalization is required that elaborates the essential data.

Based on this data, step 3, feature extraction, derives features from the plate matrix. The feature extraction scheme is responsible for calculating significant feature vectors that can be used for comparison of two plate matrices.

Finally, step 4, the recognition process, is applied to the feature vectors. In order to cope with the requirements of the parallel processing of vectors stemming from different survey points the approach uses a net-oriented recognition process. The data structure developed for recognition purposes across different survey points must provide fast insertion and retrieval of feature vectors. Therefore, a multidimensional search tree has been chosen that splits the set of feature vectors into smaller subsets regarding preselected features (such as color, etc.).

To sum up, the four steps of our image processing pipeline are:

1. Relevant frame selection
2. Plate segmentation
3. Feature extraction
4. Recognition

The following sections discuss each process in more detail.

III. RELEVANT FRAME SELECTION

The input video stream with a frame size of 512 x 512 and a frame rate of 25 frames per second is digitized using a frame grabber card resulting in a high data input rate of 6 MB per second. Dependent on the traffic density, a percentage of frames will show an empty traffic line and thus do not provide any information.

Therefore, the task of the relevant frame selection process is to filter frames of a video sequence that include a vehicle for further processing. A special requirement in the design for a suitable algorithm is the necessity for action in real time, implying that the processing time between consecutive frames must be less than 0.04 seconds. This is the upper boundary for the process duration.

As a consequence of this requirement the number of pixels involved in the decision of whether a frame contains a vehicle has to be reduced to a minimum. The reduction is achieved through the use of detection stripes spread over the frame. Their localization in the frame is not changed while processing. Because the vehicles move mainly vertically through the image, the detection stripes are arranged horizontally on the frame. We found pairs of stripes with 2 sections for the upper stripe and 3 sections for the lower stripe to be appropriate for this task (Fig. 2).

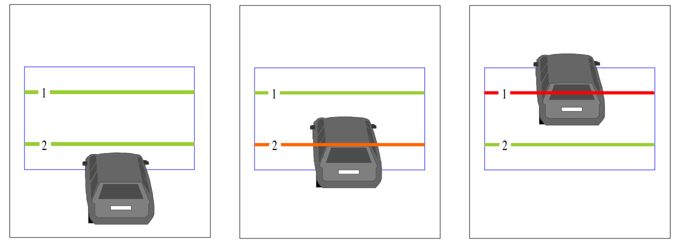

Fig. 2: Relevant frame selection method: deactivation of detection stripe 2 and simultaneous activation of stripe 1 indicates a frame of interest

The state of a detection stripe is important for the relevant frame selection. A detection stripe is said to be active if a certain percentage of its pixel are covered by a moving object. In Figure 2, we see that certain detection stripes are activated according to the position of the car in the frame. In this example a frame is selected when detection stripe 1 is active while detection stripe 2 becomes inactive. This state indicates a vehicle position where the license plate is approximately in the centre of the frame.

To find out which areas are assumed to contain a moving object, the background-difference-technique [3] is applied with the modification that it is restricted to the pixels on the detection stripes. We made further modifications to the adaptation of the background image (stripes): The entire frame information is simply transferred to the background if no detection stripes are active.

Considering the number of calculations that are necessary in processing one pixel, it turns out that a maximum of four operations are needed. These operations are visualized in the data flow graph in Figure 3, where s is the threshold which determines if a pixel is in a moving area.

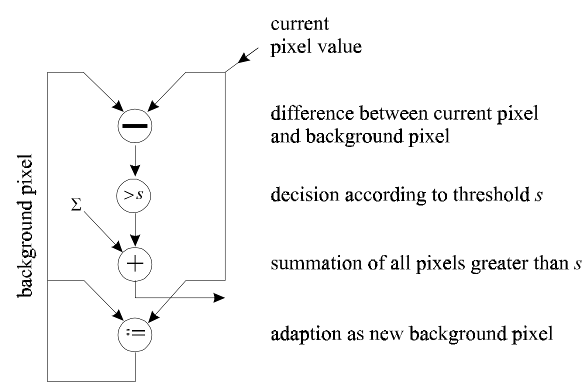

Fig. 3: Dataflow graph of pixel operations

IV. PLATE SEGMENTATION

Feature analysis of license plates requires an important preprocessing step: segmentation within frames of interest. This is performed in two steps, as outlined in Figure 4. In the first step, the image area roughly containing the license plate is cut out of the entire frame (Section IV.A). Secondly, a finer segmentation is executed in order to obtain the character field. This is the topic of Section IV.B.

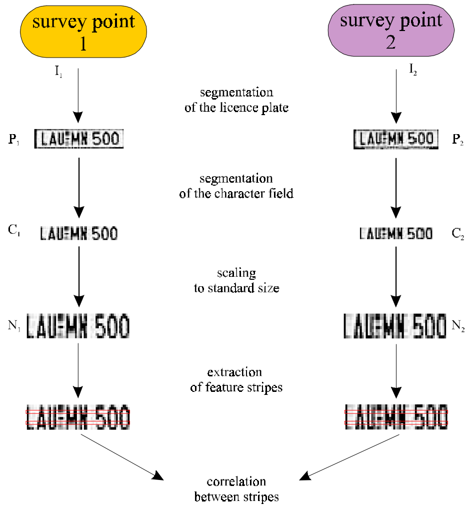

Fig. 4: The segmentation process as part of the processing pipeline

Fig. 5: Edge based license plate segmentation

A. *License Plate Segmentation*

A robust license plate segmentation is essential for all further processing steps. Its functionality, namely the determination of license plate position, can be found in a variety of video based applications such as access control systems, as well as knowledge based approaches which check image lines for white-black-white changing of character edges. Plate segments are identified when a group of image lines fulfills predefined feature thresholds.
Another approach is the adaption of artificial neural networks, which have proven their capabilities in a variety of applications in the segmentation task. The common disadvantages of segmentation methods proposed so far are:

1. They allow a more generous plate-to-frame-area relation. The projection of a license plate into an image produces a presentation which is 9 times larger than in our application;
2. They do not require real time solutions for their application.

Due to our application constraints we have developed a knowledge based solution that relies on the detection of horizontal and vertical plate edges. Accordingly, we apply a horizontal and a vertical Sobel operator [2] since it is less sensitive to noise than a Laplace filter and in addition slightly smoothes the input.

Typical plate characteristics such as white background and black border differentiate the plate's properties from other parts of the vehicle. This is highly advantageous for an edge based algorithm. The algorithm identifies the border of the license plate by analyzing the detected edges. The knowledge about the standardized size of a license plate is also used in the algorithm. Depending on the camera position and the system camera calibration the projected size of a license plate into the image plane is different for each survey point. Therefore, the algorithm is parametrized to adapt to variable location conditions at different road control stations.

Our border identifying algorithm sequentially examines the edge image E_H in Figure 5 line by line. As soon as the number of edge points in a line corresponds to a parametrized threshold, the algorithm searches for lines a certain distance away dependent on the plate size. If such a pair of lines (as lower and upper border candidates) is found, a similar search is performed for columns in edge image E_V (Figure 5). This search is restricted to the area between the two lines used as pair candidates. If this detection process is successful, a subsequent consistency check ensures that edge pixels in determined line pairs are coherent and furthermore lie between the vertical bounds.

B. *Character Field Segmentation*

As Figure 6 illustrates, the segmentation of plate images is influenced by plate properties and local illumination conditions (e.g. shadows). In order to provide comparable information for further processing, an additional segmentation which elaborates the essential information is required: namely the identification of letters inside a *character field*.

Fig. 6: Segmentation results

The segmentation starts searching for a suitable determination threshold in order to distinguish between plate background and letters. This search is realized by selecting a rectangular region in the centre of the license plate image and performing a histogram analysis of the pixel values.
After subsequent determination of the entire license plate image, the upper and lower boundary of the character field are searched. An upper boundary is found if a long coherent area of white values in a line is detected which is followed by at least three lines containing substantially shorter coherent areas of white and black values (see Figure 7). In a corresponding manner, the lower and vertical boundaries are determined.

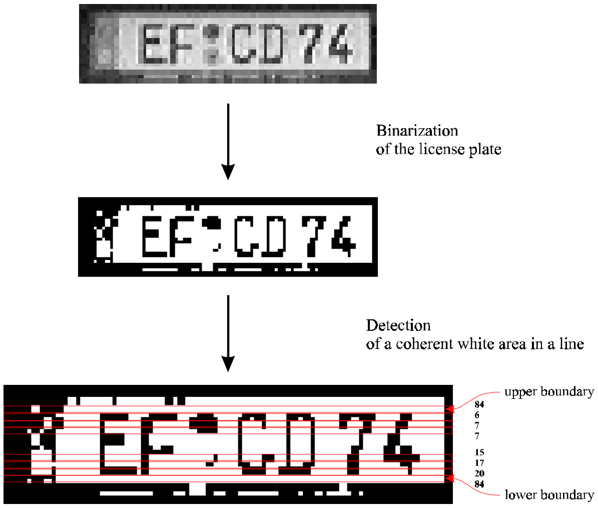

Fig. 7: Detection of the character field

V. FEATURE EXTRACTION AND RECOGNITION

In order to realize the recognition, a function *rec* is defined as a measure for the similarity of two plate image feature vectors with the signature:

rec: feature-vector x feature-vector →[0,1]

This function provides a unique value for the recognition of a vehicle at two survey points M_i and M_j. It can be used as a weighting factor in the calculation of route related traffic data. Furthermore, every survey point M is assigned a unique identification number i. For every M_i the set D_i of feature vectors is defined which is the representation of vehicles passing survey point M_i. The storage implementation of D_i uses a data structure that allows efficient insertion and retrieval of feature vectors. The data structure represents internally the geometrical topology of survey points. Thus, every topological predecessor and successor of a survey point can be determined. Additionally, stored information are distances between two consecutive survey points, and the average travel time between selected points. With these elements the recognition process is conceived as follows:

1. A feature vector $\hat{v}$ occurs at the measure point M_i
2. The set of identifiers J of all direct predecessors of M_i is determined
3. Calculate the set of candidates $V = \bigcup_{j \in J} D_j$

   It is possible to remove all elements v from V for which it is improbable that v and $\hat{v}$ are similar (for example the occurrence time difference between v and $\hat{v}$ is too small).
4. Calculate $m = \max\{rec(v, \hat{v}) | v \in V\}$
5. Identify assignment: if m is greater than the threshold S_{rec}, then $\hat{v}$ has been recognized.
6. Update the data structures and in the case of a recognition collect route information and initialize calculation of route related data.

The calculation of the feature vector for a plate matrix is the task of the feature extraction process, while the determination of m is performed by the recognition process.

For the implementation of the similarity measure *rec* we have chosen the correlation $c(\Delta x, \Delta y)$ which is mathematically defined for two real functions $f(x,y)$ and $g(x,y)$ in equation 1.

$$c(\Delta x, \Delta y) = \iint_{x,y \in \Re} f(x,y) \cdot g(x + \Delta x, y + \Delta y) \cdot \mathrm{d}x dy \quad (1)$$

Applied to our application we consider two greyscale plate images N_1 and N_2 (see Figure 4) which can be thought of as greyscale matrices or discrete functions $u(m,n)$ and $v(m,n)$. They correspond to the functions f and g in the definition above. The correlation value, which is calculated dependent on the translations, represents a measure for the similarity of the images being overlaid in a certain alignment. As a characteristic value for the similarity of two images, the maximum correlation value $\hat{c}$ is calculated over all possible alignments of the two images that can be obtained by translating the images in both directions parallel to the main axis. So far, the use of the correlation appears to be a proper solution for our similarity measure *rec*. However, $\hat{c}$ cannot be used for the comparison of different image pairs since its expressiveness is limited to a fixed choice of a single image pair. Moreover, $\hat{c}$ is not invariant regarding different illumination conditions that might occur at different survey points. The problems can be tackled by normalizing the correlation value [2].

Working with the above approach leads to a number of issues which must be resolved. First, a computational time problem needs to be solved since correlation has a computational time complexity of $O(n^4)$ considering images of size n x n. One solution is the transformation in the frequency domain. According to the Wiener-Khinchin-theorem [2], the correlation can be calculated by multiplying both Fourier transformations. Other solutions exist for the spatial domain as for example the limited consideration of possible alignments of the images or the omission of maximum value calculation. Considering the computational overhead while transforming an image into the frequency domain, the latter solutions in the spatial domain are less computationally expensive.

Secondly, the question arises whether to use black and white images instead of grey scale images. It can be shown, however, that the average correlation error is 72 times higher while using black and white images than the error when using images of 256 grey tones. Additional difficulties arise in the decision of a suitable determination threshold.

The result of our considerations is the following algorithm, which uses as input the character field of a license plate segmented as described in Section IV.B:

1. Calculate the feature vector from the image:
   a) Scale the plate image C to a standard size image N
   b) Extract n grey values from the image and write them to *field* (feature stripes)
   c) Calculate the expectancy $m = \frac{1}{n-1} \cdot \sum_{i=1}^{n} field[i]$

d) Calculate the square root from the auto correlation

$$a = \sqrt{\sum_{i=1}^{n} (field[i] - m)^2}$$

e) Create the feature vector $\mathbf{v} = (field, m, a)$

2. Calculate $rec(\mathbf{v},\mathbf{v'})$ of two feature vectors $\mathbf{v}$ and $\mathbf{v'}$

a) Calculate $r = \frac{1}{a \cdot a'} \cdot \sum_{i=1}^{n} (field[i] - m) \cdot (field[i]' - m')$

b) Calculate: $rec(v, v') = \begin{cases} 100 \cdot r & \text{if} \quad v > 0 \\ 0 & \text{otherwise} \end{cases}$

VII. RESULTS

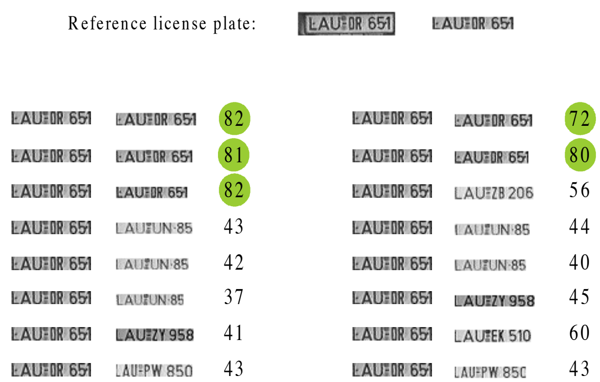

Fig. 8: Correlation based similarity measure of data set to reference license plate

The algorithms presented above have been integrated in VESUV, a comprehensive, integrated machine vision system for traffic surveillance. The VESUV system has been evaluated in a field trial on a German highway. Using 14 video cameras traffic data was recorded at a point of convergence in order to cover every traffic entrance and exit. On the base of this test data (see Fig. 8), the system's reliability as well as its time consumption were evaluated. Results are given in Table 1 and show that 60% of vehicles were recognized correctly. This recognition rate is satisfying the specified requirement, since only representatives of traffic streams are needed in order to calculate route related traffic data. The evaluation of the time consumption showed that VESUV has been able to process the video data in real time on a SUN Ultra 1 with 167 MHz.
The work presented in this paper shows promising results that will efficiently improve a new generation of traffic control systems.

	Percentage correct (individual error)	Percentage correct (superpos. of errors)
Selecting frames of interest	98%	98%
Detection of license plate	72%	70%
Correlation based identity	84%	60%

Tab. 1: Precision of net-linked data

VIII. CONCLUSION

In this paper we presented an recognition pipeline for processing video data in order to perform online route tracing by a machine vision system.
The objective of the presented approach is to minimize the computation time for the recognition by preselecting feature vectors according to the well known topology of the road network and efficiently reducing the data volume for consideration. Moreover, the time needed to compare two feature vectors is minimized by precalculation of necessary information at creation time.
Due to the use of the correlation method, our approach does not depend on optical character recognition (OCR) of the plate´s contents. Even though interpreting single characters is performed successfully in a variety of related work, the correlation approach is preferable with regard to privacy, since no identification of individuals is possible.
The outcome of our approach is the online acquisition of route related data such as travel times and tracing of traffic streams, which can highly increase the efficiency of traffic control systems.

REFERENCES

[1] Buxton, H.; Gong, S.: Advanced Visual Surveillance using Bayesian Networks. In: International Conference on Computer Vision, Cambridge, Massachusetts, (1995).
[2] Jain, A.K.: Fundamentals of Digital Image Processing, Prentice Hall, (1989)
[3] M. Kilger: A Shadow Handler in a Video-based Real-time Traffic Monitoring System. IEEE Workshop on Applications of Computer Vision, pp.1060 - 1066, (1992).
[4] Kuehne, R. et al: Loop-based travel time measurement. In: Traffic Technology International, pp. 157 - 161, (1997).
[5] Malik, J.: Robust Multiple Car Tracking with Occlusion Reasoning. In: Proc. Third European Conference on Computer Vision, Stockholm, Sweden, May 2-6, pp. 189 - 196, (1994).
[6] P.G. Michalopoulus et al.: Vehicle detection video through image processing: the Autoscope system. IEEE Transactions on Vehicle Technology, vol. 40, no. 1, (1991)