Real-time Recognition and Reidentification of Vehicles from Video Data with high Reidentification Rate

Richard WOESLER
Institute of Transport Research, German Aerospace Center
12489 Berlin, Germany

ABSTRACT

Using video cameras regions of cities can be monitored in order to extract traffic data. Stationary cameras fixed at high buildings can be used which provide data 24 hours a day. The data can be used, e.g., to optimize traffic flow by controlling traffic lights dynamically. In order to minimize the number of cameras it is useful to reidentify vehicles leaving one monitored region and afterwards entering the viewing field of a further camera. From reidentified vehicles travel times can be obtained which are relevant parameters to optimize traffic control. In the present text, a method to reidentify vehicles based on extraction of 3-d-prototype vehicle models and color extraction from the top plane of vehicles is described. Shadows and light reflections on wet street are corrected, and therefore, the high recognition accuracy is achieved which is necessary to find the top plane of the vehicles. Due to the 3-d-model based analysis the cameras can be placed in a broad region of viewing angles. The algorithms are suitable for real-time applications. First results from video data of two cameras are presented which show a high reidentification rate with no false reidentification hypothesis.

Keywords: traffic data extraction, video data, real-time, reidentification of vehicles, vehicle length extraction, vehicle color extraction, 3-d-model based analysis.

1. INTRODUCTION

Traffic data extraction from video data is of great value, e.g., in order to optimize traffic flow by dynamic traffic light control and dynamic control of electronic speed signals, and in order to obtain detailed information where improvements of infrastructure are to be made. In co-operation with universities and a number of smaller and larger companies from Berlin and Brandenburg [1,2], we are interested to monitor, e.g., regions of the city of Berlin with its usually about 40 thousand cars traveling at the same time. Many systems exist concerning the question of how to extract traffic data from video signals [3-6]. A powerful method is 3-d model based analysis [7,8], where 3-d models of vehicle prototypes are compared with extracted object hypotheses. This method will be of increasing interest in the future as computer power further increases [8-11].

Another approach is region based tracking. The idea is to identify a connected region in the image associated with each vehicle and then track it over time using a cross-correlation measure [12,13]. A dual to the region based approach is tracking based on active contour models, or 'snakes'. The idea is to have a representation of the bounding contour of the object and keeping updating it dynamically [14].

All these approaches exhibit a problem of handling partially occluded vehicles, because such vehicles are usually recognized as vehicle clusters and not as single vehicles. A method to solve this problem is to track vehicle sub-features [15]. In our approach we also extract sub-features, i.e. we extract corners of the vehicles, therefore, our approach is appropriate to expand it to solve this problem, too.

In order to obtain reidentification data about vehicles from video data, alternatively it is possible to extract the information of the number plates. However, for larger regions this is expensive, because a large number of cameras is necessary with high resolution and in appropriate viewing angles.

Instead, we perform the recognition of vehicles from rather arbitrary viewing angles by using a 3-d-model based approach. We show that our approach provides high accuracy, e.g., concerning the extracted lengths of vehicles. Due to this high accuracy, our algorithm is able to extract a part behind the middle part of the top plane of an object. From this part, appropriate color values are extracted, converted to HSV (hue, saturation, value) color space and used as parameters for reidentification of objects [16]. The reason for extracting data not from the middle part is that there are frequently openings in the front part of the top plane of the vehicles.

We present concepts and results about fast recognition of vehicles with high accuracy, and about fast reidentification of vehicles for videos of two cameras.

2. CONCEPT FOR FAST VEHICLE RECOGNITION

Movement detection and merging of movement regions

Movement regions on streets can be detected using difference images with a certain threshold. In order to suppress slight shadows, a renormalization with respect to brightness can be used, i.e. color values RGB are converted to r=R/(R+G+B), g=G/(R+G+B), b=1-r-g. This yields

$$q_p(t)=H\Big(\left|r_p(t)-r_p(t-\tau)\right|+\left|g_p(t)-g_p(t-\tau)\right|+$$

$$+\left|b_p(t)-b_p(t-\tau)\right|-\theta\Big)\ . \qquad (1)$$

Here, p denotes the index of the pixel, and the function H(z) is zero for z<=0 and one for z>0. The threshold is θ, which depends on the camera, and the time difference between two images is chosen to be τ=1/6 s, which is suitable for city traffic for vehicle recognition. The regions with $q_p(t)=1$ are called movement regions.

Regions of interest (ROIs) are defined within traffic areas like streets as follows. 3-d-world coordinates are obtained for the streets using GPS. A middle line within each street can be defined in the image by defining in 1 m distances points and connecting them. Analogously, lines which correspond approximately to the boundary of the streets can be defined.

Movement regions are merged together when they are close to each other.

Definition of object hypothesis

Two movement regions successive in time in roughly the same region are candidates for vehicles. Around both regions a rectangular frame is defined with one direction along the street and the other direction perpendicular to the street. Both rectangles overlap and the largest rectangle which belongs to both rectangles is defined to be the uncorrected rectangle of the object hypothesis.

The rectangular frame around the object hypothesis sometimes contains the shadow of the object and sometimes contains reflections especially of vehicle lights like vehicle beacons on wet surface, see Fig. 1.

Such errors can be corrected by comparing regions of smoothed edge images of prototypes of a corresponding corner of 3-d-models of vehicles with the corresponding region of a smoothed edge image of the original image within the uncorrected rectangle of the object hypothesis, see Fig. 1 (b).

The region which corresponds best to the corresponding corner is chosen, and the corresponding corner position is defined to be the (corrected) corner of the (corrected) rectangle of the object hypothesis.

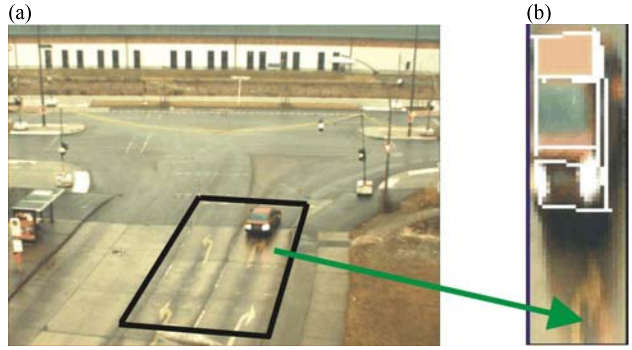

Fig. 1. (a) The black rectangle on the road shows the region where currently vehicles are recognized by the software. The light of the right vehicle beacon is reflected on the wet road into the camera. (b) The extracted (uncorrected) rectangle around the object parallels the black rectangle of Fig. 1 (a). The reflection of the light of the vehicle beacon is contained in the extracted (uncorrected) frame (arrow). Therefore, the rectangle is much larger than the image of the car, i.e. it is too large. Within (b) the result of the corner detection is illustrated, too. The lower right aft corner of the vehicle is detected by the software. Afterwards, a new rectangle is computed where the lower right corner of the (uncorrected) rectangle is corrected to the place of the detected corner. Then the software projects a 3-d-model of a car into the new rectangle. This is shown as the white grid model. It can be seen that the corner has been detected at the right position, and therefore, the reflection error is corrected by the software. Further explanations about the algorithms are given in the main text.

Testing vehicle hypotheses and definition of hypotheses of physical dimensions of objects

The described (corrected) rectangles of the vehicle hypotheses are tested by comparing them with regions which are covered by projections of 3-d-models of vehicles (see Fig. 2) projected into the scene in such a way that they occupy the same rectangle.

This procedure is simplified to enable fast vehicle recognition as follows: Only at the points of a 2m times 2m grid these projections are computed, and they are computed before the online processing is done. Further, the 3-d-models contain only about twenty 3-d-points, corners of the vehicle. The displacement to the exact position and the scaling to the extracted rectangle is done only for the corners. By this method, the length, width, and height of the scaled model are obtained (Fig. 1 (b)). These values have to lie within a certain region of possible physical dimensions of the corresponding vehicle prototype, else the prototype is skipped for this object hypothesis. By this method, a car can be distinguished by a truck. Then, the extracted physical dimensions are the hypotheses for the corresponding physical dimensions of the object.

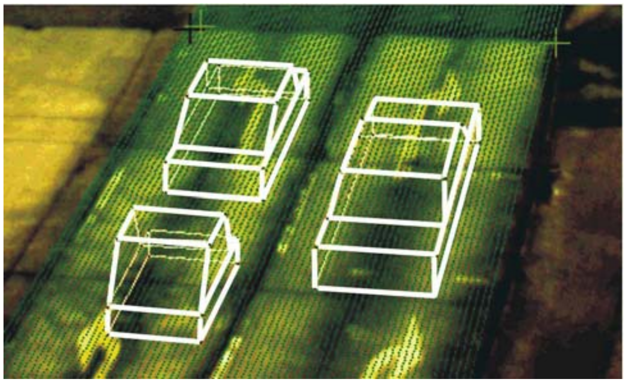

Fig. 2. Some 3-d grid models of vehicles projected into the image. Such models are used to compare them with edge images of real extracted object hypotheses.

Extraction of hypotheses of vehicle colors

The position of the projection of a point which is 30cm behind the middle of the top plane of the 3-d vehicle model is known. In case that the resolution of the cameras is not worse than about 0.1m on the corresponding part of the road plane in the direction perpendicular to the vector which connects the street point with the camera, we find with our method, that this extracted point lies well within the real top plane of the vehicle (Fig. 1 (b)). Therefore, it makes sense to use the color values extracted around this position as color values of this plane of the vehicle.

However, reflections from the sky disturb these values. Therefore, a vertical polarizing filter is used which corrects the values, which works especially well close to the Brewster angle.

These color values are converted to hue-saturation-value (HSV) color space, because the value hue yields for a large part of the color space a stable parameter which encodes the tone of the color of a vehicle rather precisely. Roughly speaking, the parameter hue is invariant against changes of brightness and remains invariant when adding white light to the values RGB.

3. RESULTS OF VEHICLE RECOGNITION

An RGB video of traffic in Berlin-Adlershof at the crossing Rudower Chaussee with Wegedornstreet with 50 vehicles turning from the Wegedornstreet to the right into the Rudower Chaussee is recorded with 6 Hz frame rate with vertical polarizing filter.

The region of the road which is analyzed is the region in front of the traffic light in Fig. 1 (a), ranging from the traffic light back to the lower boundary of the image. 100% of the vehicles are recognized by the algorithm.

The extracted hypotheses of the physical dimensions of the vehicles are compared with the actual dimensions of the vehicles [17]. The lengths have a standard deviation from the actual values of 0.4 m. The widths and heights have a standard deviation of 0.2 m.

These results demonstrate the high accuracy of the software, because the objects are seen from behind and, therefore, it is difficult to extract the correct lengths. The resolution of the image was about 0.1m.

For the color values extracted from extracted hypotheses of the uppermost vehicle plane (Fig. 1 (b)) the empirical standard deviations for each single vehicle are computed for multiple measurements of the vehicle. The mean of these standard deviations were 3% in H (where H is a value between 0 to 360 degrees), 3% in S (which is a value between 0 and 1), and 2% in V (which is a value between 0 and 1).

The complete computations have to be conducted for every third image, i.e. with a rate of 2 Hz. The computation lasts 0.2 s for each three images on a 2 GHz usual personal computer. Therefore, it is faster than real-time. In future it is planned to use also a field programmable gate array (FPGA), which will further increase computation speed.

4. CONCEPT FOR FAST VEHICLE REIDENTIFICATION

The reidentification is conducted with the extracted vehicle parameters described above after the vehicles have been tracked within the viewing field of each camera. Tracking is performed according to the place and velocity of an object and an object of the following time step according to the assumption of constant velocity allowing for some deviations regarding place and velocity. The tracking algorithm did not yield any error in the videos analyzed so far.

The number of parameters is much lower than, e.g., comparing the whole vehicle image data, therefore, fast reidentification can be achieved.

Reidentification can be done by testing the hypothesis that the differences of the parameter values are stochastically for two extracted objects. In case that some approximations are made, especially normally distributed extracted parameter values, and statistical independence of the parameters, theoretical equations exist. We get

$$U(m,k) = \left[\sqrt{2}\cdot a(n,\alpha)\right]^{-1} \cdot \sqrt{\sum_{i=1}^{n}\left(\frac{\mathrm{d}(i,m,k)}{\mathrm{b}(i)\,\mathrm{s}(i,m,k)}\right)^2} \qquad (2)$$

where U(m,k) is an error function for the extracted and tracked object m of the first camera, and k that of the second camera, respectively. The parameters we use are for i=1 the color tone H, for i=2 the color saturation S, for i=3 the brightness V, for i=4 the vehicle length, and for i=5 the parameter time. n=5 is the number of parameters taken into account. The parameter α is the level of significance. The function $a(n,\alpha)$ is related to the incomplete Gamma function. With $\alpha=0.05$ we get $a(5,0.05)\approx 3.33$. We define H1(m) to be the parameter value hue H of the tracked object m in camera 1. We define H2(k) to be the parameter value H of the tracked object k in camera 2. Analogous definitions are made for the other parameters saturation S1(m), brightness V1(m). length L1(m), time t1(m) etc.. The function d(i,m,k) is the absolute difference of the parameter values of the two objects for i=1,2,3,4. For i=5 it is

$$d(5,m,k) = |t2(k) - t1(m) - t\_R| . \quad (3)$$

The value t_R is an estimation of the travel time which is computed according to the velocity and the estimated acceleration of the objects as follows. For vehicles slower than the recommended velocity v_recomm = 50km/h an acceleration towards the recommended velocity and afterwards constant velocity are assumed. For vehicles faster than v_recomm constant velocity is assumed. The following auxiliary variables are defined. For v1(m)<v_recomm the time to reach v_recomm for constant acceleration is

$$t\_change = t1(m)+[v\_recomm - v1(m)] /a\_usual , \quad (4)$$

where a_usual is set to 2.75 m/(s·s). The travel time for constant acceleration would be

$$t\_R\_aux = \left(\sqrt{v1(m)^2 + 2 \cdot a\_usual \cdot (y2(k) - y1(m))} - v1(m)\right) \cdot (a\_usual)^{-1} \quad (5)$$

where y1(m), and y2(k) respectively, are the positions of a corner of the vehicle along the street. The algorithm for computation of t_R is:

if [v1(*m*) >= v_recomm] then t_R = [y2(*k*)-y1(*m*)]/v1(*m*) ,

else if [t_change < t1(*m*) + t_R_aux] then

$$t\_R = \{y2(k) - y1(m) - [t\_change - t1(m)] \cdot v1(m) + - 0.5 \cdot a\_usual \cdot [t\_change - t1(m)]^2\} / v\_recomm + + (v\_recomm - v1(m)) / a\_usual$$

else

$$t\_R = t\_R\_aux . \quad (6)$$

In Eq. (2) the functions s(i,m,k) are estimations of empirical standard deviations. They can be optimized, e.g., using neural networks. It is also possible to use empirical knowledge to estimate values for these parameters. As an ansatz we use the following equations.

$$Max\_V = Max (V1(m) , V2(k)) , \quad (7)$$

$$Max\_V\_part = (1.2/255)^2 / (Max\_V + 1/255)^2 . \quad (8)$$

The Max_V_part equation is designed for an 8bit RGB matrix camera, therefore, the value 2^8-1 = 255 is used. Further we define

$$Max\_S = Max (S1(m) , S2(k)) , \quad (9)$$

$$s(1,m,k) = [10 + 250 \cdot (d(3,m,k))^2 + 70 \cdot (d(2,m,k))^2 + +5/(Max\_S + 0.005) + 200 \cdot Max\_V\_part] \text{ degree} , \quad (10)$$

$$s(2,m,k) = 0.035 + 0.5 \cdot d(3,m,k) + Max\_V\_part , \quad (11)$$

$$s(3,m,k) = Min (0.06 + 0.0001 \cdot |t1(m)-t2(k)|, 0.15) , \quad (12)$$

$$s(4,m,k) = 60 \text{ cm} , \quad (13)$$

$$s(5,m,k) = t\_R / 3 . \quad (14)$$

The equations mirror the knowledge that for larger deviations of the time larger variations usually occur in the parameter values, because brightness usually changes more when the time difference is larger, and that for large parameter value differences the values obtain a larger standard deviation, because the values are then closer to the boundary values, at which the camera usually shows larger standard deviations for the parameters, since the CCD camera sensor which was used was not designed for high quality color results.

In Eq. (2) the values b(i) are weighting parameters. In case they all were one a usual hypothesis testing would be conducted. However, different parameters have different relevance, e.g., hue H is more relevant than the brightness V, because the latter may vary rather much between the two regions monitored by the cameras for the same object.

The distance between the regions of the two cameras was 60m. For this distance the time slot in which objects are taken into account for reidentification is defined to range from t1(m)+t_R/1.4 up to t1(m)+t_R·1.8.

5. RESULTS OF VEHICLE REIDENTIFICATION

The regions monitored by the two cameras are shown in Fig. 3. Two lanes are analyzed for cloudy weather conditions. Camera 1 is upstream. Camera 2 is downstream. The reidentification is complicated by the pedestrian crossing and by the crossing of vehicles in the images of camera 2. Further some vehicles slow down and leave the analyzed road at the crossing. Two videos were used. The first video was the training data to

(a) (b)

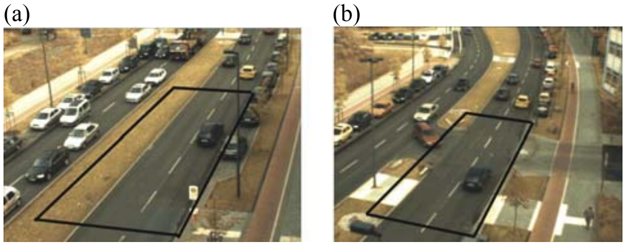

Fig. 3. Rudower Chaussee in Adlershof, Berlin, Germany. (a) An image of camera 1 is shown. The black rectangle on the road indicates the region where vehicles are analyzed by the software. (b) An image of camera 2 is shown with the region which is analyzed.

optimize parameters. The second video was the test data to verify the results. The first video of camera 1 was about 8min long with 39 cars. 38 of the cars are recognized by the algorithm. The discrimination between cars and larger vehicles worked without an error. The data base which was used does not work perfect: the data base does not receive all extracted data sets. For the first video only 36 of the 38 obtained data sets from vehicles were stored in the data base. The tracking algorithm takes the data from the data base. 100% of the 36 vehicles are tracked correctly. The weighting parameters in Eq. (2) were

$$b(1) = 1.8, b(2) = 2.8, b(3) = 4, b(4) = 4, b(5) = 0.26. \quad (15)$$

For U<1.3 the interim hypothesis of possibly same vehicle is defined. In case that the second smallest U-value for a vehicle is larger than U_1=0.1 plus the smallest U-value the final hypothesis is defined that the objects are the same. In case one wants to decrease the rate of false reidentification hypotheses, which is already zero for all data analyzed so far, further, one can increase the value of U_1. The lane changing is restricted so that it does not exceed 1.9m, because the distance between the images of the cameras is small which restricts lane changing. Further reidentification hypotheses are made when there are two vehicles A and B which may correspond to a third vehicle C, and there is a fourth vehicle D which can only correspond to A or to B and not to A and to B according to the U-values. Then D is identified with A, or B respectively, and C with B, or A respectively. The reidentification algorithm yields a reidentification rate of 100% of the 36 vehicles. No false reidentification hypothesis occurred.

The test data is a video of about 17.5 min length with 117 cars in the images of camera 1. From these cars for camera 1 there are 102 (=87%) recognized and passed into the data base. For camera 2 there are 105 (=90%) recognized, and 101 (=86%) are passed into the data base. Further, the number of cars which occur in the data base from both cameras was 89. The reidentification yields 78 reidentified cars. No false reidentification hypothesis occurs. The corresponding rate of reidentification is 78/89=88%. The reidentification rate of the total number of 117 cars is 78/117=67%.

We remark that due to technical limitations the frame rate of the images was not optimal. The mean rate was 4.4 Hz. Most of the images had a rate of 5 Hz, the rest of them a rate of 2.5 Hz. The algorithm is designed to work optimally for about 6 Hz, therefore, 2.5 Hz is too low for good recognition results. The number of recognition errors due to 2.5 Hz frame rate was 7. Therefore, 117-7=110 cars should be recognized by the algorithm which means a rate of recognized cars passed into the data base of 102/110=93%. The corresponding reidentification rate is 78/110=71%.

The tracking and reidentification processes, and the computation of travel times and mean velocities, including reading data from database and writing results into a file for the two times about 17.5 min data with 117 cars lasts about 2.3 s on a common 1.7 GHz personal computer. Therefore, the reidentification scheme is well suited for real-time applications.

Another analysis was done to compare the data of cars coming from Wegedornstreet, Fig. 1 (a), turning to the right into Rudower Chaussee, Fig. 3. In addition to the first ROI, which was placed as shown in Fig. 1 (a) in the Wegedornstreet, a second ROI was placed in the bend. For reidentification only color HSV and length were used, i.e. n=4, compare Eq. (2). Reidentification could be done successfully [18]. This demonstrated that for different viewing angles of the cameras reidentification can be conducted due to the 3-d-model based analysis which allows finding the top plane of the vehicles for different viewing angles.

6. CONCLUSION

With the concept presented it is possible to extract traffic data with high accuracy in real-time from video data. Due to an explicit vehicle corner detection (Fig. 1) errors like reflections on wet surface can be corrected so that the accuracy is sufficient for reidentification of vehicles.

The concept to extract the colors of the top plane of a car, by using 3-d-model based vehicle recognition, yields the advantage that traffic data from rather arbitrary viewing directions can be obtained. Therefore, the number of cameras which are necessary to monitor a given region is usually lower than for systems with rather restricted viewing angles, and, therefore, the system is comparatively cheap. Additionally, the possibility of reidentification of vehicles reduces the number of cameras further, because with reidentification it is not necessary to monitor directly the complete region.

The next step is to obtain and to analyze data from cameras with large distance between the monitored regions. The vehicle color and vehicle length will be weighted the stronger the larger the distance is, i.e. in Eq. (15) b(1), b(2), b(4), and 1/b(5) will be chosen smaller. Especially cars with characteristic saturated colors are optimal for reidentification [18]. Further we want to optimize reidentification of vehicles by extending the software by algorithms for platoon matching [19] and to reidentify sets of vehicles. Further the initial estimation of the travel time, Eq. (6), shall be done by a faster than real-time microscopic simulation [20] of the detected vehicles.

7. ACKNOWLEDGEMENTS

Financial support by the German government department BMBF is gratefully acknowledged. The author thanks Axel Berres, Martin Schlicker, and Ines Ernst (all same Institute of Transportation Research as the author) for the implementation of parts of the recognition algorithms, and he thanks Bjoern Appel (same Institute) for taking part in the evaluation of the data.

8. REFERENCES

[1] see http://www.dlr.de/vf/forschung/projekte/ois
[2] see http://www.fav.de
[3] S. Eggiman, J. Tschirren, "Benchmarking Bildverarbeitungssysteme für den Straßenverkehr. Bureau des autoroutes Fribourg", Suisse, 1996.
[4] N. Zheng, and J. D. Crisman, "Vehicle matching using color," Proc. of the **IEEE Conf. on Intelligent Transport Systems**, 1998, pp. 206–211.
[5] H. H. Topp, J. Riel, and A. Lagemann, "Fortentwicklung eines Verfahrens zur Auswertung und Analyse automatisch erhobener Daten von Verkehrsströmen an Straßenquerschnitten", **Forschung Straßenbau und Straßenverkehrs-technik**, BVBW, Germany, Vol. 823, 2001.
[6] V. Kastrinaki, M. Zervakis, K. Kalaitzakis, "A survey of video processing techniques for traffic applications", **Image and Vision Computing**, Vol. 21, 2003, pp. 359–381.
[7] D. Koller, "Detektion, Verfolgung und Klassifikation bewegter Objekte in monokularen Bildfolgen am Beispiel von Straßenverkehrsszenen", **DISKI**, Vol. 13, Germany, infix–Verlag, 1992.
[8] D. Koller, K. Daniilidis, and H. H. Nagel, "Model based object tracking in monocular image sequences of road traffic scenes", **International Journal of Computer Vision**, Vol. 10, 1993, pp. 257–281.
[9] K. D. Baker and G. D. Sullivan, "Performance assessment of model-based tracking", Proc. of the **IEEE–Workshop on Applications of Computer Vision**, Palm Springs, CA, 1992, pp. 28–35.
[10] G. D. Sullivan, "Visual interpretation of known objects in constrained scenes", **Phil. Trans. Roy. Soc. B**, Vol. 337, 1992, pp. 361–370.
[11] M. Haag and H.H. Nagel, "Incremental recognition of traffic situations from video image sequences", **Image and Vision Computing**, Vol. 18, 2000, pp. 137–153.
[12] K. P. Karmann and A. von Brandt, "Moving object recognition using an adaptive background memory", **Time-Varying Image Processing and Moving Object Recognition**, V. Casppellini, editor, 2. Elsevier, Amsterdam, The Netherlands, 1990.
[13] M. Kilger, "A shadow handler in a video-based real-time traffic monitoring system", **IEEE Workshop on Applications of Computer Vision**, Palm Springs, CA, 1992, pp. 1060–1066.
[14] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russel, "Towards robust automatic traffic scene analysis in real-time", **ICPR**, Israel, 1994.
[15] B. Coifman, D. Beymer, P. P. McLaughlan, and J. Malik, "A real-time computer vision system for vehicle tracking and tracking surveillance", **Transportation Research Part C**, Vol. 6, 1998, pp. 271–288.
[16] J. D. Foley, A. v. Dam, S. K. Feiner, J. F. Hughes, **Computer Grafics**, Addison Wesley, 1995.
[17] **ADAC Autokatalog 2001**, ISBN 3-6264-0855-1, Germany, ADAC Verlag, 2001.
[18] R. Woesler, "Fast extraction of traffic data and reidentification of vehicles from video data", **Proc. IEEE 6th Internat. Conf. On Intell. Transportation Systems**, Shanghai, China, 2003, pp. 774-778.
[19] B. Coifman, and M. Cassidy, "Vehicle reidentification and travel time measurement on congested freeways," **Transportation Research Part A**, Vol. 36, 2002, pp. 899–917.
[20] K. Nagel, P. Wagner, and R. Woesler, "Still Flowing: Approaches to Traffic Flow and Traffic Jam Modeling", **Operations Research**, Vol. 51, No. 5, 2003, pp. 681-710.