TRAJECTORY-BASED SCENE DESCRIPTION AND CLASSIFICATION BY ANALYTICAL FUNCTIONS

D. Pfeiffer [a], R. Reulke [b*]

[a] Akeleiweg 46D, 12487 Berlin, Germany - david@dltmail.de
[b] Humboldt-University of Berlin, Computer Science Department, Computer Vision, Rudower Chaussee, Berlin, Germany - reulke@informatik.hu-berlin.de

Commission III, WG III/4 and III/5

KEY WORDS: Object recognition, Scene Analysis, statistical and deterministic trajectory models

ABSTRACT:

Video image detection systems (VIDS) provide an opportunity to analyse complex traffic scenes that are captured by stationary video cameras. Our work concentrates on the derivation of traffic relevant parameters from vehicle trajectories. This paper examines different procedures for the description of vehicle trajectories using analytical functions. Derived conical sections (circles, ellipses and hyperboles) as well as straight lines are particularly suitable for this task. Thus, it is possible to describe a suitable trajectory by a maximum of five parameters. A classification algorithm uses these parameters and takes decisions on the turning behaviour of vehicles.

A model based approach is following. The a-priori knowledge about the scene (here prejudged and verified vehicle trajectories) is the only required input into this system. One confines himself here to straight lines, circles, ellipses and hyperboles. Other common functions (such as clothoids) are discussed and the choice of the function is being justified.

1. INTRODUCTION

1.1 Motivation

Traffic management is based on an exact knowledge of the traffic situation. Therefore, traffic monitoring at roads and intersections is an essential prerequisite. Inductive loops and microwave radar systems are the most common detection and surveillance systems to measure traffic flow on public roads.

VIDS that operate with real time image processing techniques became more attractive during the last 15 years (Michalopoulos 1991), (Wigan 1992), (Setchell et al. 2001), (Kastrinaki et al. 2003). Traditional traffic parameters like presence, vehicle length, speed as well as time gap between two vehicles and vehicle classification (Wei et al. 1996) can be determined. In contrast to other sensors, the use of local cameras makes a two-dimensional observation possible and thus can determine new traffic parameters like congestion length, source-destination matrices, blockage or accidents and therefore support the estimation of travel times. Multi-camera systems extend some limitations of single camera systems (e.g. occlusions, reliability) and enlarge the observation area (Reulke et al. 2008a).

We proposed a framework that autonomously detects atypical objects, behavior or situations even in crowded and complex situations (Reulke et al. 2008b). Extracted object data and object trajectories from multiple sensors have to be fused. An abstract situational description of the observed scene is obtained from the derived trajectories. The first step in describing a traffic scene is to ascertain the normal situation by statistical means. In addition, semantic interpretation is also derived from statistical information (such as direction and speed). Deviations of the inferred statistics are interpreted as atypical events, and therefore can be used to detect and prevent dangerous situations. These options allow the detection of sudden changes as well as atypical or threatening events in the scene. Atypical or threatening events are generally defined as deviations from the normal scene behavior or have to be defined by a rule based scheme. Red light runners and incident detection systems are an example for a self-evident road traffic application.

The trajectories of street vehicles are smooth and homogeneous over a large scale. Therefore, a mathematical description by elementary functions is appropriate for these trajectories. Thus, dramatic reductions of the bandwidths are achieved for a full scene transmission. The basic step to determine the driver intentions is to fit the trajectories to the analytical functions.

This paper is organized as follows: After an overview of situation analysis and atypical event detection the approach is introduced. Then, an example installation is described and its results are presented. The mathematical fundamentals of the adaptation of formerly derived trajectories of turning vehicles by hyperbolas, ellipsoids, spheres and straight lines are sketched. The derived information is very comprehensive but compact and permits downcast to other representations like source destination matrices. The paper closes with a summary and an outlook.

1.2 Situation Analysis and Atypical Event Detection

Scene description and automatic atypical event detection are issues of increasing importance and an interesting topic in many scientific, technical or military fields where complex situations (i.e. scenes containing many objects and interac-

---

* Corresponding author.

tions) are observed and evaluated. A common aim is to describe the observed data and to detect atypical or threatening events.

Other areas of situation analysis besides driver assistance (Reichardt 1995) may include traffic situation representation, surveillance applications (Beynon et al. 2003), sport video analysis or even customer tracking for marketing analysis (Leykin et al. 2005).

(Kumar et al. 2005) developed a rule-based framework for behavior and activity detection in traffic videos obtained from stationary video cameras. For behavior recognition, interactions between two or more mobile targets as well as between targets and stationary objects in the environment have been considered. The approach is based on sets of pre-defined behavior scenarios, which need to be analyzed in different contexts.

(Yung et al. 2001) demonstrate a novel method for automatic red light runner detection. It extracts the state of the traffic lights and vehicle motions from video recordings.

1.3 Image and Trajectory Processing

The cameras deployed cover overlaid or adjacent observation areas. With it, the same road user can be observed using different cameras from different view positions and angles. The traffic objects in the image data can be detected using image processing methods.

The image coordinates of these objects are converted to a common world coordinate system in order to enable the tracking and fusion of the detected objects of the respective observation area. High precision in coordinate transformation of the image into the object space is required to avoid mis-identification of the same objects that were derived from different camera positions. Therefore, an exact calibration (interior orientation) as well as knowledge of the position and view direction (exterior orientation) of the camera is necessary.

Since the camera positions are given in absolute geographical coordinates, the detected objects are also provided in world coordinates.

The approach is subdivided into the following steps. Firstly, all moving objects have to be extracted from each frame of the video sequences. Secondly, these traffic objects have to be projected onto a geo-referenced world plane. Afterwards, these objects are tracked and associated to trajectories. One can now utilize the derived information to assess comprehensive traffic parameters and to characterize trajectories of individual traffic participants.

1.4 Scenario

The scenario has been tested at the intersection Rudower Chaussee / Wegedornstrasse, Berlin (Germany) by camera observation using three cameras mounted at a corner building at approximately 18 meters height. The observed area has an extent of about 100x100 m and contains a T-section. Figure 1 shows example trajectories derived from images taken from three different positions. The background image is an ortho-photo, derived from airborne images.
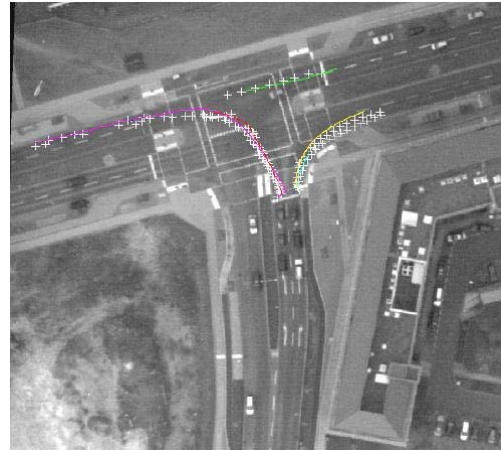

Figure 1. Orthophoto with example trajectories

The aim is the description of the trajectories by functions with a limited number of parameters. Source destination matrices could be determined at these crossroads through such parameters without any further effort. A classification approach shall be used here.

2. PROCESSING APPROACH

2.1 Video Acquisition and Object Detection

In order to receive reliable and reproducible results, only compact digital industrial cameras with standard interfaces and protocols (e.g. IEEE1394, Ethernet) are deployed.

Different image processing libraries or programs (e.g. OpenCV or HALCON) are available to extract moving objects from an image sequence. We used a special algorithm for background estimation, which adapts to the variable background and extracts the desired objects. The dedicated image coordinates as well as additional parameters like size and area were computed for each extracted traffic object.

2.2 Sensor Orientation

The existing tracking concept is based on extracted objects, which are geo-referenced to a world coordinate system. This concept allows the integration or fusion of additional data sources. The transformation between image and world coordinates is based on collinearity equations. The Z-component in world coordinates is deduced by appointing a dedicated ground plane. An alternative is the use of a height profile. Additionally needed input parameters are the interior and exterior orientation of the camera. For the interior orientation (principal point, focal length and additional camera distortion) of the cameras the 10 parameter Brown distortion model (Brown 1971) was used. The parameters are being determined by a bundle block adjustment.

Calculating the exterior orientation of a camera (location of the projection centre and view direction) in a well known world coordinate system is based on previously GPS measured ground control points (GCPs). The accuracy of the points is better than 5 cm in position and hight. The orientation is deduced through these coordinates using DLT and the spatial resection algorithm (Luhmann 2006).

2.3 Tracking and Trajectories

The aim of tracking is to map observations of measured objects to existing trajectories and to update the state vector describing those objects, e.g. position or shape. The tracking is carried out using a Kalman-filter approach.
The basic idea is to transfer supplementary information concerning the state into the filter approach in addition to the measurement. This forecast of the measuring results (prediction) is derived from earlier results of the filter. Consequently, this approach is recursive.

The initialization of the state-vector is conducted from two consecutive images. The association of a measurement to an evaluated track is a statistical based decision-making process. Errors are related to clutter, object aggregation and splitting. The decision criteria minimize the rejection probability.

The coordinate projection mentioned in the last paragraph and the tracking process provides the possibility to fuse data acquired from different sensors. The algorithm is independent of the sensor as long as the data is referenced in a joint coordinate system and they share the same time frame.

The resulting trajectories are then used for different applications e.g. for the derivation of traffic parameters (TP).

2.4 Trajectory analysis

A deterministic description method for trajectories shall be introduced below. The functional descriptions for these trajectories should be as simple as possible and permit a straightforward interpretation. Linear movements will be described by simple straight lines.

Numerous suggestions of possible functions for curve tracks by functional dependencies have been made in the literature. Clothoid (Liscano et al. 1989) or G2-Splines (Forbes 1989) are curves whose bend depends on the arc length. Alternatively, closed functions like B-Splines, Cartesian polynomials fifth degree or Polarsplines (Nelson 1989) can be used as well. A common approach to approximate vehicle-based trajectories is to employ clothoids. Those functions derived from the fresnel integral are highly non linear. They are fundamental in road and railroad construction. Due to urban constraints the tracks of intersections and curves cannot follow the curve of a clothoid whose shape is regarded as a trajectory that is especially comfortable to drive. Because there are only partial approximations of clothoids, they do not fit into the set of elementary functions that shall be regarded in this work. Moreover, the given trajectory has to be subdivided into parts in order to apply a clothoidal approximation. (Anderson et al. 1979) have proposed a description of tracks by hyperbolas. The great advantage is that the derived parameters clarify directly geometric connections and permit a categorization and derivation of important features of the trajectories. A hyperbola is able to replicate straight lines as well as turning trajectories.

The hyperbola fit serves as an example and is described next. The approach is based on least-square fitting of geometric elements. The equation for a hyperbola with semi-major axis parallel to the x-axis and semi-minor axis b parallel to the y-axis is given by

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{1}$$

The parametric equation is given by

$$x = a \cdot \sec(t) \quad y = b \cdot \tan(t) \tag{2}$$

Commonly the hyperbola is rotated and shifted:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} x' - x_m \\ y' - y_m \end{pmatrix} \tag{3}$$

Wherein m_x, m_y are the centre coordinates, the angle φ is the bearing of the semi-major axis. The implicit form of the hyperbola can be written as a general polynomial of second degree:

$$a_1' \cdot x'^2 + a_2' \cdot x'y' + a_3' \cdot y'^2 + a_4' \cdot x' + a_5' \cdot y' = 1 \tag{4}$$

With

$$a_1' = \frac{a_1}{a_4} \quad a_2' = \frac{a_2}{a_4} \quad a_3' = \frac{a_3}{a_4}$$
$$a_4' = -\frac{2 \cdot x_m a_1 + y_m a_2}{a_4} = -\left(2 \cdot x_m a_1' + y_m a_2'\right)$$
$$a_5' = -\frac{x_m a_2 + 2 \cdot y_m a_3}{a_4} = -\left(x_m a_2' + 2 \cdot y_m a_3'\right) \tag{5}$$

and

$$a_1 = \frac{\cos^2\varphi}{a^2} - \frac{\sin^2\varphi}{b^2}$$
$$a_2 = 2 \cdot \sin\varphi \cdot \cos\varphi \cdot \left(\frac{1}{a^2} + \frac{1}{b^2}\right)$$
$$a_3 = \frac{\sin^2\varphi}{a^2} - \frac{\cos^2\varphi}{b^2} \tag{6}$$
$$a_4 = 1 - \left(x_m^2 a_1 + x_m y_m a_2 + y_m^2 a_3\right)$$

The following equations describe the conversion of the implicit to the hyperbola parametric form:

- Bearing of the semi-major axis

$$\varphi = a\tan\frac{a_2'}{a_1' - a_3'} \bigg/ 2 \tag{7}$$

- Center coordinates

$$x_m = -\frac{2 \cdot a_4' a_3' - a_5' a_2'}{4 \cdot a_1' a_3' - a_2'^2} \tag{8}$$
$$y_m = -\frac{2 \cdot a_1' a_5' - a_4' a_2'}{4 \cdot a_1' a_3' - a_2'^2}$$

43

- Semi-major axis

$$a^2 = \frac{1 - \left(a_4' \cdot \frac{x_m}{2} + a_5' \cdot \frac{y_m}{2} \right)}{a_1' \cdot \cos^2 \varphi + a_2' \cdot \sin \varphi \cdot \cos \varphi + a_3' \cdot \sin^2 \varphi} \qquad (9)$$

$$b^2 = -\frac{1 - \left(a_4' \cdot \frac{x_m}{2} + a_5' \cdot \frac{y_m}{2} \right)}{a_3 \cdot \cos^2 \varphi - a_2 \cdot \sin \varphi \cdot \cos \varphi + a_1 \cdot \sin^2 \varphi}$$

The parameter determination is based on the number of observations n, which are related to the functional model. The number of observations n has to be greater than the number of unknown parameters.

$$a_1' \cdot x_i'^2 + a_2' \cdot x_i' y_i' + a_3' \cdot y_i'^2 + a_4' \cdot x_i' + a_5' \cdot y_i' = 1 \qquad (10)$$

or

$$\begin{bmatrix} x_0'^2 & x_0' y_0' & y_0'^2 & x_0' & y_0' \\ x_1'^2 & x_1' y_1' & y_1'^2 & x_1' & y_1' \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n-1}'^2 & x_{n-1}' y_{n-1}' & y_{n-1}'^2 & x_{n-1}' & y_{n-1}' \end{bmatrix} \cdot \begin{bmatrix} a_1' \\ a_2' \\ a_3' \\ a_4' \\ a_5' \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \qquad (11)$$

This can also be written as follows:

$$\underline{l} = \underline{\underline{A}} \cdot \underline{x} \qquad (12)$$

The observation vector $\underline{l}$ is replaced by the measured observation and a small residuum v. Therefore, the unknown vector x is replaced by the estimates with the result:

$$\hat{\underline{x}}_u = \left(\underline{\underline{A}}^T_{u,n} \cdot \underline{\underline{P}}_{n,n} \cdot \underline{\underline{A}}_{n,u} \right)^{-1} \cdot \underline{\underline{A}}^T_{u,n} \cdot \underline{\underline{P}}_{n,n} \cdot \underline{l}_n \qquad (13)$$

This result is known as a least-square adjustment, based on the L_2 norm. This approach is not able to decide between hyperbola and ellipsoid. (Fitzgibbon et al. 1996) and (Fitzgibbon et al. 1999) describe an attempt for the inclusion of additional conditions by integration of a constraint matrix. Hence it is possible to reduce the resulting solution space so that the type of the object function (ellipse, hyperbole) can be steered. (Harlow et al. 2001) and (Harker et al. 2008) enlarge Fitzgibbon's approach by decomposition of the Scattermatrix in the square, linear and constant part. The parameter estimate becomes equivalent to the eigenvalue problem. This is a direct solution method. The approach determines an ellipse as well as two hyperboles. Figure 2 shows examples for the hyperbola and ellipse fit.
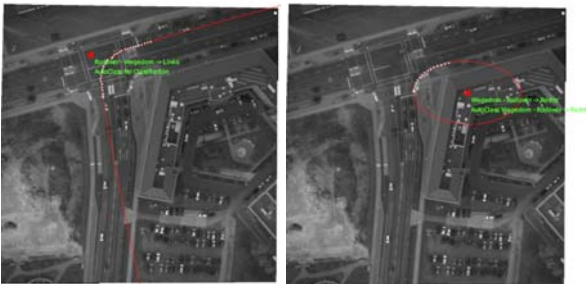

Figure 2. Example fits for two tracks and classification

3. CLASSIFICATION

3.1 Class definition

The traffic objects are identified within the image data and trajectories are derived from it. The trajectories are fitted to curves and their parameters are classified with the corresponding functions. For the classification the same data set is used for all function classes.

A part of the data set is used to train a classifier which intends a class assignment for the trajectory with the parameters. The other part serves for the verification.
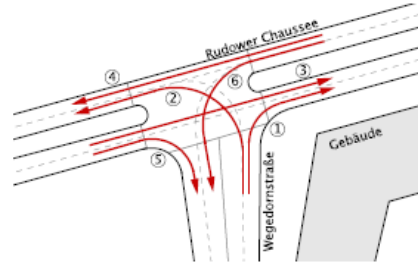

Figure 3. Visualisation of the different traffic lanes (classes) of the scene

Seven classes were defined based on the scene (figure 3) and the traffic lanes:

No	From	To	Class	Abbreviation
1	Wegedorn	Rudower	right-turn	WRR
2	Wegedorn	Rudower	left-turn	WRL
3	Rudower	Rudower	east-direction	RO
4	Rudower	Rudower	west-direction	RW
5	Rudower	Wegedorn	right-turn	RWR
6	Rudower	Wegedorn	left-turn	RWL
7	No class membership			No_Class

Table 1. Class definition for the observed scene

The used data set consists of 414 trajectories. Trajectories which are part of the classification process need to have a minimal length of 10 m and a minimal number of points of at least 6 points. The class No_Class consist of trajectories of pedestrians, bicyclists and erroneous tracks caused by errors in image processing and tracking. It is inadmissible that two driving directions are assigned to one trajectory. Relying on the shape only, opposite directions is merely to distinguish, since their functional parameters are similar. To achieve the distinction the approximate same path of the trajectory and the fitted function is determined. With this, the direction of the trajectory can be determined as an additional feature. Hence it direction can be distinguished between close lanes trajectories with opposite directions.

3.2 Classification method

A classifier determines the class affiliation with the characteristic of item-specific features. These features are represented as a vector in a multidimensional feature space. The features correspond to the parameters which have been determined by the approximation.

A rectangle classifier and a modified k nearest neighbours (KNN-) classifier are used. The result of the classification shall be unambiguously.

k-nearest neighbours algorithm (KNN) is a method for classifying objects based on closest training examples in the feature space.

The rectangle classification (also cuboid classification) is a distribution free, nonparametric and supervised classification method (see figure 4).
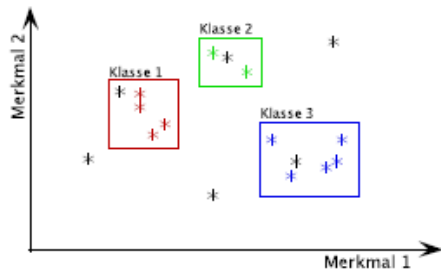

Figure 4. A simple rectangle classification in a 2D feature space

The KNN classification needs a training data set. It is a non-parametric method for the estimate of probability densities. The operation of the classifier is steered by k (number of regarded neighbours, a free selectable parameter) and δ (used metric). Figure 5 shows the approach.

Figure 5. Visualization of the KNN classification. The k=7 nearest neighbour are used. The object g is assigned to the class B

The metric δ defines the reliable determination of the distances to adjacent elements. The result of the classification depends substantially on the density of the learning set and the choice of the metric. Here the Mahalanobis distance was used.

4. RESULTS

	To-tal	NC	WRR	WRL	RO	RW	RWR	RWL
Ref	414	62	117	59	72	26	33	54
Circ	414	62	119	59	68	21	34	51
Elli	410	51	117	58	72	28	34	50
Hyp	410	51	117	58	72	28	34	50
Str	413	50	125	56	70	28	35	49

Table 2. Summary of complete occurrence and the class occurrence of different trajectory types. Ref – reference, Circ – circle, Elli – ellipse, Hyp – hyperbola, Str - straight lines

A data set of 414 different trajectories (Total) has been processed using different functions within the test data set. A total of 62 trajectories could not be classified (NC). A summary is given in table 2.

The results shall be represented in greater detail by the hyperboles in the following.

4.1 Hyperbola

Figure 6 shows examples of the approximation of hyperboles.

Figure 6. Approximation of hyperboles

In addition to the parameters of the conical sections the direction of motion was uses for the classification. Figure 7 shows the plot of the rotational angle φ (X) and the delta in degrees (ϕ) where the trajectory adapts to the hyperbola:
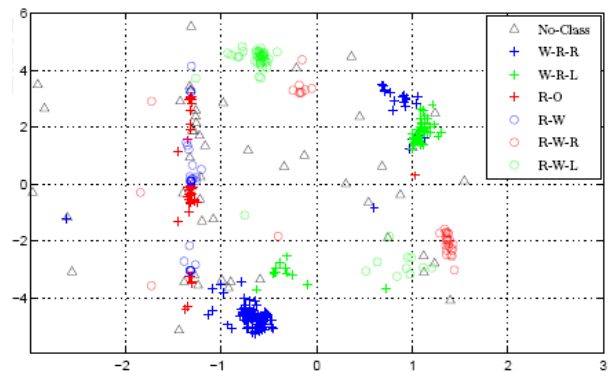

Figure 7. Classification results

Class	Cuboid Classifier	KNN-Classifier
Total	92.9%	97.8%
No_Class	84.3%	98.0%
WRR	95.7%	99.1%
WRL	93.1%	99.0%
RO	97.2%	99.5%
RW	96.4%	96.4%
RWR	85.3%	91.2%
WRL	92.00%	94.0%

Table 3. Comparison of results for hyperbolas fits achieved by Coboid and KNN-Classifier

Figure 7 shows a clear separation of the feature space. A high classification rate is achieved by both elementary classifiers (see table 3).

5. CONCLUSION AND OUTLOOK

Table 3 affirms a high reliability on these elementary functions, with respect to the used basic classification methods. Mistakes within the classification mostly reside due to scene behaviour that occurs fairly rare (e.g. car turning at the intersection) or is not modelled by the underlying functions (e.g. pedestrians or cyclists crossing in very custom patterns). The shown approaches have been tested and verified in a real-time environment with a multi-camera system.

The system shall to automatically observe the traffic on crossroads in future. For example source-destination dependences can be determined with that.

6. REFERENCES

Anderson, B. and J. Moor (1979). Optimal filtering. Enlewood Cliffs, New Jersey, Prentice-Hall, Inc.

Beynon, M. D., D. J. V. Hook, et al. (2003). Detecting Abandoned Packages in a Multi-Camera Video Surveillance System. IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'03).

Brown, D. C. (1971). "Close range camera calibration." Photogrammetric Engineering 37(8): 12.

Fitzgibbon, A., M. Pilu, et al. (1999). "Direct least square fitting of ellipses." Pattern Analysis and Machine Intelligence, IEEE Transactions on 21(5): 476-480.

Fitzgibbon, A. W., M. Pilu, et al. (1996). Direct least squares fitting of ellipses. Pattern Recognition, 1996., Proceedings of the 13th International Conference on.

Forbes, A. B. (1989). Least-squares best fit geometric elements, National Physical Laboratory of Great Britain.

Harker, M., P. O'Leary, et al. (2008). "Direct type-specific conic fitting and eigenvalue bias correction." Image Vision Comput. 26(3): 372-381.

Harlow, C. and Y. Wang (2001). "Automated Accident Detection System." Transportation Research Record 1746(-1): 90-93.

Kastrinaki, V., M. Zervakis, et al. (2003). "A survey of video processing techniques for traffic applications." Image and Vision Computing 21(4): 359-381.

Kumar, P., S. Ranganath, et al. (2005). "Framework for real-time behavior interpretation from traffic video." Intelligent Transportation Systems, IEEE Transactions on 6(1): 43-53.

Leykin , A. and M. Tuceryan (2005). A Vision System for Automated Customer Tracking for Marketing Analysis: Low Level Feature Extraction. International Workshop on Human Activity Recognition and Modelling. Oxford, UK.

Liscano, R. and D.Green (1989). Design and implementation of a trajectory generator for an indoor mobile robot. Proceedings of the IEEE/RJS International Conference on Intelligent Robots and Systems. Tsukuba, Japan: 380–385.

Luhmann, T. (2006). "Close Range Photogrammetry: Principles, Methods and Applications."

Michalopoulos, P. G. (1991). "Vehicle detection video through image processing: the Autoscope system." Vehicular Technology, IEEE Transactions on 40(1): 21-29.

Nelson, W. L. (1989). "Continuous steering-function control of robot carts." IEEE Transactions on Industrial Electronics 36(3): 330–337.

Reichardt, D. (1995). A Real-time Approach to Traffic Situation Representation from Image Processing Data. Intelligent Vehicles Symposium. Detroit, MI, USA.

Reulke, R., S. Bauer, et al. (2008a). Multi-Camera Detection and Multi-Target Tracking. VISAPP 2008. Funchal, Madeira (Portugal).

Reulke, R., S. Bauer, et al. (2008b). Situation Analysis and Atypical Event Detection with Multiple Cameras and Multi-Object Tracking. Robot Vision. Auckland (New Zeeland), Springer-Verlag Berlin Heidelberg.

Setchell, C. and E. L. Dagless (2001). "Vision-based road-traffic monitoring sensor." Vision, Image and Signal Processing, IEE Proceedings - 148(1): 78-84.

Wei, C.-H., C.-C. Chang, et al. (1996). "Vehicle Classification Using Advanced Technologies." Transportation Research Record 1551(-1): 45-50.

Wigan, M. R. (1992). "Image-Processing Techniques Applied to Road Problems." Journal of Transportation Engineering 118(1): 21.

Yung, N. H. C. and A. H. S. Lai (2001). "An effective video analysis method for detecting red light runners." Vehicular Technology, IEEE Transactions on 50(4): 1074-1084.