Multiple Vehicle 3D Tracking Using an Unscented Kalman Filter

Daniel Ponsa[†], Antonio López[†], Joan Serrat[†], Felipe Lumbreras[†] and Thorsten Graf[‡]

***Abstract*— This article describes a system to track vehicles on images taken from a mobile platform. The objective is to determine the position and velocity of vehicles ahead of the mobile platform, in order to make possible the prediction of their position in future instants of time. This problem is addressed by modeling a 3D dynamic system, where both the acquisition platform and the tracked vehicles are represented in a state vector. From measurements obtained in every frame, this state vector is re-estimated using an Unscented Kalman Filter, instead of the Extended Kalman Filter used in previous works. Assuming that vehicles progress on a flat surface, a novel model of their dynamics is proposed, which explicitly considers constraints on the velocity. With respect to previous approaches, this model improves tracking reliability, since the estimation of unfeasible states is avoided. Experiments on real sequences display promising results, although a more systematic evaluation of the system should be done.**

I. INTRODUCTION

The research in Computer Vision applied to intelligent transportation systems is mainly devoted to provide them with situational awareness, either to ascertain the state of their driver and passengers or specially to characterize elements on its external environment. Indeed, most of the applications required by the automobile industry [1] (lane departure warning, automatic cruise control, autonomous stop & go driving and lane change assistance), rely on an accurate perception of the own vehicle surroundings. In this paper, the task of characterizing vehicles on the road ahead of a mobile platform is considered, using images provided by a single monochrome camera. The objective is estimating not only the relative 3D position of vehicles, but also their velocity. This allows to predict future vehicle locations to, for instance, warn about unsafe manoeuvres.

Many references in the literature deal with vehicle detection and tracking using a monocular system. However, just a few of them have been thoroughly tested on long and complex sequences. The pioneer work in [2] describes a vehicle detection and tracking method, which was deeply tested on public roads in the Vamp and VITA II prototypes. Once vehicles are detected, their relative 3D position and velocity with respect to the mobile platform is estimated using an Extended Kalman Filter, following the strategy first described in [3].

A similar approach is adopted in [4][5], tested in the Navlab 6 prototype. It proposes a more complete description of the problem, taking into consideration not only the 3D state of tracked vehicles, but also the pose and velocity of the platform holding the camera. A different proposal, claimed robust to different lightening and meteorological conditions, is the one in [6]. The system is based on a color camera, and vehicles, once detected, are tracked in image coordinates by a region matching algorithm. No 3D model is used during tracking, and the only 3D information that is inferred from tracked 2D regions is a rough estimation of the distance of vehicles with respect to the mobile platform.

To develop the applications demanded by the automobile industry, reliable 3D information about tracked vehicles is needed. Therefore, it is natural to address this problem using a 3D model based approach. This allows not only to estimate directly the information of interest, but also to make a more coherent and robust analysis of sequences. Specifically, the dynamics of vehicles can be more accurately modeled and visual occlusions among them can be explicitly considered. This is the approach adopted in this work. The principal contributions of this paper are the use of constrained models to describe the dynamics of vehicles, and the application of the Unscented Kalman Filter, instead of the Extended Kalman Filter, to estimate the state of vehicles.

The outline of the paper is as follows. Next section describes the acquisition system used in this work, specifying a projective model to relate acquired images with the observed world. Section III continues describing a 3D model of the system which, like in [4], maintains in its state vector both the mobile platform and the tracked vehicles. An accurate model of the state dynamics is proposed, which accounts explicitly for the range of feasible velocities in vehicles. Section IV details the process to obtain measurements of vehicles from images, which are then used in the Unscented Kalman Filter described in section V to reestimate the pose of tracked vehicles. Section VI briefly describes a global vehicle detection–tracking framework that uses the proposed tracking scheme, and presents some examples of its performance. Section VII draws some conclusions and describe some work in developed.

II. ACQUISITION SYSTEM SETUP

The acquisition system consist of a monochrome camera, mounted on the rear-view mirror support of a vehicle (in the following, the *host* vehicle) facing the road ahead with a slight inclination. The camera has a $1/3$ Inch CMOS sensor, and a fixed premounted optics with focal distance around 7 mm. Images are digitized at a resolution of 640×480 pixels.

This research has been partially funded by Spanish MEC project TRA2004-06702/AUT. We thank to Marc Seguer and Francisco Sánchez at SEAT for their collaboration in the capture of the video sequences.

([†]) Computer Vision Center and Dept. d'Informàtica, Universitat Autònoma de Barcelona. {daniel, antonio}@cvc.uab.es

([‡]) Volkswagen AG Group Research, Electronics.

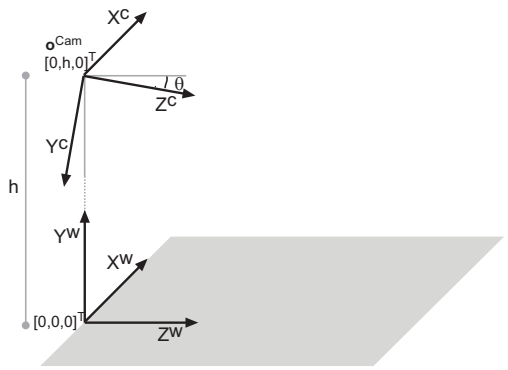

Fig. 1. Camera coordinate system relative to the host coordinate system

A. Camera Model

To model the acquisition system, first a host coordinate system is defined relative to the camera position, placed externally on the road that sustains the host vehicle. This coordinate system gives the reference to relate the camera with the 3D world that acquires (figure 1). Using this convention, the point of view from where the images are captured (i.e., the camera extrinsic parameters) it is completely defined by the camera origin $\mathbf{o}^{Cam} = [0, h, 0]^T$ and the pitch angle θ describing its inclination with respect to the road surface.

The projection of the 3D world onto 2D images (i.e. the camera intrinsic parameters) is modeled as a pin-hole camera with zero-skew [7]. This requires identifying the effective camera focal length on the image X and Y direction (f^x, f^y) an its center of projection (x^0, y^0). This has been done calibrating the camera with the software provided in [8]. Using homogeneous coordinates, the image projection of a 3D coordinate $[x, y, z]^T$ on images is defined by

$$s \begin{bmatrix} x^c \\ y^c \\ 1 \end{bmatrix} = \mathbf{A}\,\mathbf{S}\,\mathbf{R}\,\mathbf{T} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

where s is a scale factor; $\mathbf{A}$ is the camera intrinsic matrix; $\mathbf{S}$ is a reflection matrix to model the different sign of the Y axe of the camera reference system; $(\mathbf{R}, \mathbf{T})$ specify the rotation and translation which locate the camera coordinate system with respect to the host coordinate system. For the situation in figure 1, this parameters are

$$\mathbf{A} = \begin{bmatrix} f^x & 0 & x^0 \\ 0 & f^y & y^0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix},$$

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -h \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

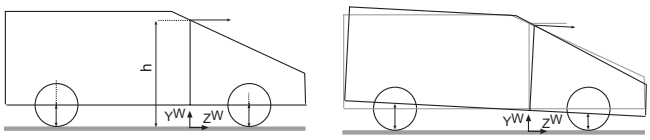

Fig. 2. Simplified vehicle suspension system. Left) Neutral position. Right) Camera position and pitch variation

From (1), the image coordinates $[x^c, y^c]^T$ corresponding to a world coordinate $[x, y, z]^T$ are determined by

$$x^c = x^0 + \frac{f^x\, x}{z\, \cos(\theta) + (-h + y)\, \sin(\theta)}, \tag{2}$$

$$y^c = y^0 + \frac{f^y\, ((h - y)\cos(\theta) + z \sin(\theta))}{z\, \cos(\theta) + (-h + y)\, \sin(\theta)}. \tag{3}$$

B. Extrinsic Parameters Variation

An important fact that can not be ignored is that the position of the camera with respect to the host coordinate system can vary significantly in every frame. Indeed, the effect of the suspension system of the vehicle alters the camera extrinsic parameters (see Figure 2). As long as the host vehicle used does not provide information about its suspension system, uncertainty in the camera extrinsic parameters has to be explicitly considered. Evaluating how the projection of points on the road vary due to changes in the extrinsic parameters, it is found that the parameter with a more significant impact is θ. Ignoring variations on $\mathbf{o}^{Cam}$ provokes a small error negligible for the goals of this paper. For this reason, this paper only considers variations on θ.

III. 3D MODEL

Given a sequence, the variation observed between frames is mainly devoted to the change of position of:

- the observed vehicles (the tracking *targets*).
- the host vehicle (which changes the camera viewpoint).

The objective in this paper is to ascertain from images the state of all this vehicles. This problem is addressed using the classical approach in estimation theory, which requires the definition of :

- a system state $\mathbf{x}$, maintaining the information of vehicles to be estimated.
- a dynamical model, defining the expected evolution along time of parameters in $\mathbf{x}$.
- an observation model relating $\mathbf{x}$ with observations of vehicles extracted from acquired images.

These elements are then combined in the procedure illustrated in figure 3. Given $\hat{\mathbf{x}}_{t|t}$ (the more likely state at instant t given observations up to this same instant), the system model is used to predict its expected state at the next instant, $\hat{\mathbf{x}}_{t+1|t}$. Then, the observation model is applied to synthesize the measurements $\hat{\mathbf{y}}_{t+1|t}$ that should be observed at frame $t+1$, given that $\hat{\mathbf{x}}_{t+1|t}$ is correct. This predicted measurements are then compared with real measurements $\mathbf{y}_{t+1}$ extracted from the current image. The disparity between them conforms the *innovation* term, which is used to update the system state to $\hat{\mathbf{x}}_{t+1|t+1}$. This is scarcely the procedure carried out by

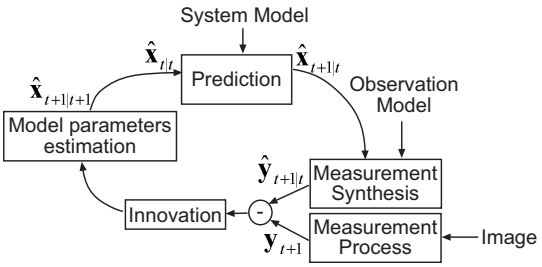

Fig. 3. Classical model based estimation cycle.

estimation methods. In fact, this methods do not estimate just a single state value, but the distribution of the more likely states given the evidence in the observations.

A very important point in this iterative scheme is the initialization of the system state at the first iteration. For space reasons, in this paper this initialization step has been omitted, putting more stress in the system model, the extraction of vehicle measurements from images, and the estimation algorithm used.

A. System state

The positional variation of vehicles along time is maintained in the state vector $\mathbf{x}$. Assuming that the road conforms to a planar surface, the movement of a vehicle is modeled by a velocity vector parallel to this plane. Extending the proposal in [4] to multiple vehicle tracking, $\mathbf{x}$ maintains

- The velocity describing the change of position and orientation of the host vehicle between frames $(v^h, d\psi^h)$.
- The position (x^i, z^i) and orientation (yaw angle) ψ^i of each tracked vehicle i with respect to the host, and its corresponding velocity $(v^i, d\psi^i)$.

So, if n vehicles are being tracked, $\mathbf{x}$ corresponds to

$$\mathbf{x} = \{(v^h, d\psi^h), (x^i, z^i, \psi^i, v^i, d\psi^i)_{i=1:n}\}$$

Figure 4 illustrates graphically the meaning of the state parameters. Their values are represented using two different reference frames.

- $(v, d\psi)$ of host and tracked vehicles are expressed in terms of a fixed global 3D reference frame. They respectively maintain their forward velocity and yaw rate.
- Position (x^i, z^i) and orientation ψ^i of targets are specified in the host coordinate system (see section II-A).

The advantage of using different reference frames is that each parameters is described in a coordinate system where its dynamics are simpler and better modeled. Considering $(v, d\psi)$ under a global reference frame allow to make realistic assumptions about its dynamics, which otherwise can not be done if a relative coordinate frame is used. For example, that their values remain constant between frames. On the other hand, it is better to describe the pose of targets (x^i, z^i, ψ^i) relatively to the host coordinate system. This is really the information of interest, and in that way it is avoided to estimate the position and orientation of the host vehicle with respect to a global coordinate system.

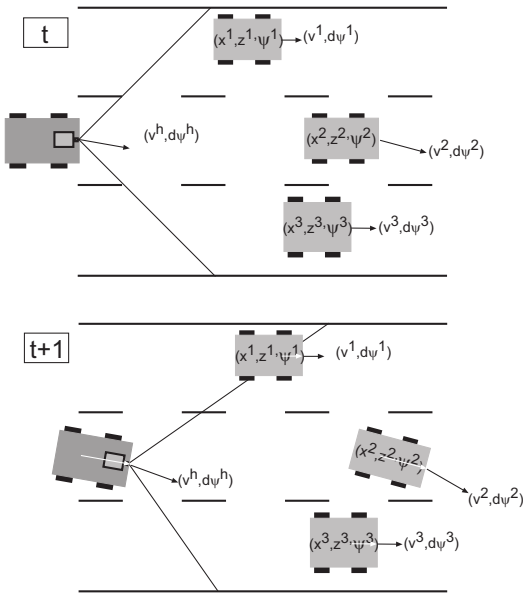

Fig. 4. Sketch showing system parameters at consecutive time instants.

B. System Model

The state evolution along time is expressed using a discrete-time nonlinear dynamic system, given by

$$\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t) + \mathbf{n}^{\mathbf{x}}{}_t \ , \tag{4}$$

where $\mathbf{F}$ specifies the expected evolution of $\mathbf{x}_t$t and $\mathbf{n}^{\mathbf{x}}{}_t$ is an stochastic noise term following a normal distribution, that accounts for the inaccuracies of the model in $\mathbf{F}$. The different parameters in $\mathbf{x}$ evolve using different expressions. In the following they are concretely specified.

1) $(v, d\psi)$ dynamics : The behavior along time of vehicles depends strongly on the road type where they move. In this paper vehicles driving on highways or A roads are considered. In this context the ground plane assumption holds most of the time, and the behavior of vehicles is usually regular and smooth. So, it is expected that vehicles maintain approximately its velocities $(v, d\psi)$ along time, experimenting just smooth changes frame–by–frame. In many works this behavior is modeled using a first-order auto-regressive process known as Brownian motion model. With respect to the evolution of v [1], this model corresponds to

$$v_{t+1} = v_t + bn_t^v \ , \tag{5}$$

where n_t^v is a random value following a normal distribution $\mathcal{N}(0, 1)$, and b a scale factor that magnifies/diminishes the disturbance that n_t^v represents. The value of b depends on the feasible vehicle acceleration considered between frames. Expression (5) represents well the expected velocity behavior, but it has the important drawback that its values evolve unconstrainedly. Indeed, at long term, v_t may took values that a vehicle is unable to reach. For the vehicle tracking problem, a priori knowledge is available delimiting

[1] Equivalent expressions and reasonings are derived for $d\psi$.

the range of forward velocities that are feasible. To consider this constraint in the evolution of v_t, this paper proposes the use of a constrained Brownian motion model (CBM, see [9] for details), expressed by

$$v_{t+1} = av_t + bn_t^v \ . \tag{6}$$

a is a scalar defined by $a^2 = 1 - \epsilon$, where $0 < \epsilon \ll 1$. At short term, the evolution of v_t is very similar to the one obtained using expression (5). However, at long term it can be seen that the values generated are constrained inside the Gaussian envelope of a normal distribution with parameters

$$\mathcal{N}(0, \frac{1}{\epsilon}bb^T) \ . \tag{7}$$

Using that, a model of dynamics is established that

- (C1) constrains v_t in a range $[-l^v, l^v]$.
- (C2) forces the term bn_t^v to take an average magnitude value equivalent to m^v (the more likely vehicle acceleration expected).

This is achieved in the following way. It is well known that a random variable with normal distribution $\mathcal{N}(0, \sigma^2)$ is constrained with a 99.73% probability in the range $[-3\sigma, 3\sigma]$. From this property, the Gaussian envelope in (7) establishes consequently a range of more likely v_t values, that it can be adjusted to have desired bound values. Thus, a CBM that fulfills constraint C1 requires that

$$l^v = 3\frac{b}{\sqrt{\epsilon}}. \tag{8}$$

Another property of random variables with normal distribution $\mathcal{N}(0, \sigma^2)$ is that the expectation of their absolute value corresponds to $\sigma\sqrt{2/\pi}$. From this fact, a CBM fulfilling C2 requires a b value given by

$$b = \sqrt{\frac{\pi}{2}}m^v. \tag{9}$$

Combining (8) and (9), the value of ϵ generating the desired CBM corresponds to:

$$\epsilon = \frac{3\pi}{2}\left(\frac{m^v}{l^v}\right)^2 \ .$$

Figure 5 displays a simulation of the evolution of v_t using the usual approach based on (5), and the one proposed using (6). Under same initial conditions and equivalent disturbances, the new proposal maintains the velocity in a given range of feasible values.

2) (x^i, z^i, ψ^i) dynamics: Estimating the relative position of targets at each frame requires to

- Reexpress their pose in the previous frame relatively to the new position of the host coordinate frame.
- Update their pose given their current velocities.

For the i-th target, this process (illustrated in figure 6) reduces to:

$$\begin{aligned} \psi_{t+1}^i &= \psi_t^i + d\psi_{t+1}^i - d\psi_{t+1}^h \ , \\ \begin{bmatrix} x_{t+1}^i \\ z_{t+1}^i \end{bmatrix} &= \mathbf{R}(d\psi_{t+1}^h)\begin{bmatrix} x_t^i \\ z_t^i \end{bmatrix} - \begin{bmatrix} 0 \\ v_{t+1}^h \end{bmatrix} + \\ &\quad + \mathbf{R}(\psi_{t+1}^i)\begin{bmatrix} 0 \\ v_{t+1}^i \end{bmatrix} \ , \end{aligned}$$

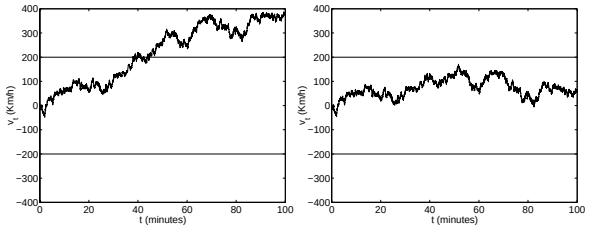

Fig. 5. Simulation of the evolution of a v_t using an unconstrained (left) and constrained (right) Brownian motion model. Horizontal lines mark the range of feasible velocities considered.

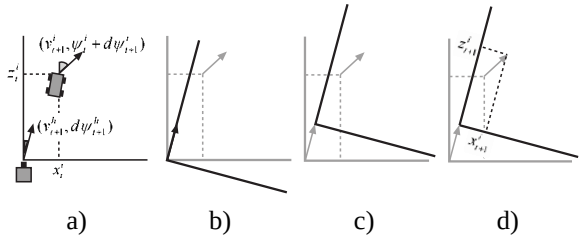

Fig. 6. Prediction of a target new pose. a) Previous host and target state, and current velocities. b) & c) New position of the host coordinate frame. d) Target translation.

where $\mathbf{R}$ is a rotation matrix with respect to the Y axis.

C. Observation Model

The observation model relates the system state $\mathbf{x}_t$ with the information extracted from images $\mathbf{y}_t$. Considering that the observation process is disturbed by additive Gaussian noise $\mathbf{n}^{\mathbf{y}}{}_t$, this is modeled as

$$\mathbf{y}_t = \mathbf{H}(\mathbf{x}_t) + \mathbf{n}^{\mathbf{y}}{}_t \ . \tag{10}$$

In this paper, function $\mathbf{H}$ synthesizes the following procedure. Given $\mathbf{x}_t$, for each target the 3D coordinates (in host coordinates) of a cuboid are established, delimiting the 3D bounding cube where each target is supposed to be found. The vertexes of this cuboids are parameterized from the pose (x^i, z^i, ψ^i) of targets, and its 3D dimensions $[w^i, h^i, l^i]^T$ established when initialized. Then, the front or rear face of each cuboid (the one visible from the camera point of view) is projected on image coordinates using equations (2) and (3), determining the 2D image region where the front/back of a vehicle is seen. This 2D region is what is expected to be extracted from images by the measurement process described in section IV. Notice that the projection equations used to obtain the 2D regions are conditioned on θ, and therefore first its value at each frame has to be ascertained in some way. An estimation of this value is obtained at the end of the measurement extraction procedure.

IV. MEASUREMENT EXTRACTION

To extract from images the rectangular regions showing the front/back face of vehicles (the one visible from the camera viewpoint), it is proceeded in the following way: Given the state prediction $\hat{\mathbf{x}}_{t+1|t}$, the value of its corresponding observations $\hat{\mathbf{y}}_{t+1|t}$ is computed for $\theta \in [\theta_{min}, \theta_{max}]$, which is the range of values that θ can take according to the possible

Fig. 7. Measurement Process. Dashed lines show the projection of the back face of the vehicle cuboid, for the bound values of the θ range. Solid line show the measurement obtained, which determines θ_{best} for this frame.

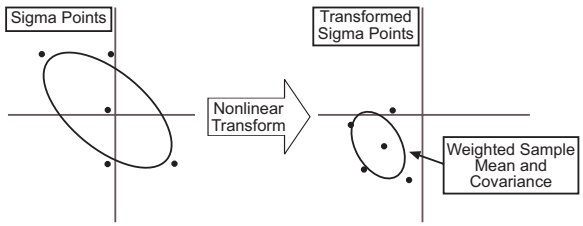

Fig. 8. the Unscented Transform mechanism to estimate the mean and covariance of a Gaussian distribution, when is propagated nonlinearly.

states of the host suspension system. Joining the different observations computed, a group of 2D regions where to look for vehicles is obtained. This set is then augmented by the slight translation and scaling of its members, and analyzed by a vehicle classifier. The classifier judges if the sub-image in each 2D region matches the appearance of a vehicle model learned previously, generating as result a list of *positive* 2D regions. The positive regions of each target in $\hat{\mathbf{x}}_{t+1|t}$ are then clustered together, setting finally $\mathbf{y}_{t+1}$ with the average region of each cluster. Once $\mathbf{y}_{t+1}$ has been obtained, an iterative search procedure is started (see figure 7), to determine the camera pitch θ_{best} that generates the projection $\hat{\mathbf{y}}_{t+1|t}$ closer to it. This θ_{best} value is considered the current camera pitch and determines the observation model $\mathbf{H}$ used in (10) .

The vehicle detector used in this procedure is inspired on the proposal in [10]. Given a labeled training set of 2D regions showing vehicles and non–vehicles, an schema based on the Adaboost algorithm identifies the visual features more appropriate to detect vehicles, and combines them to construct a classifier. The visual features considered are a family of Haar-like basis function, because they can be computed very efficiently, allowing the use of the designed classifier in real–time applications.

V. SYSTEM ESTIMATION

In this paper the Unscented Kalman Filter(UKF) [11] is used to estimate the distribution of $\mathbf{x}_t$ along time, given observations in $\mathbf{y}_t$. This algorithm generalizes the Kalman Filter (KF) to nonlinear systems, avoiding the linearisation steps required by the Extended Kalman Filter (EKF). It is easy to implement, and its performance is superior to that of the EKF. Moreover, it also can deal with non-Gaussian noise inputs (although this is not the case of the system modeled). Compared to more general estimation methods as Particle Filters, requires a lower computational effort.

Algorithms based on the KF represent the distribution of the system state at instant t by means of a Gaussian distribution $\mathcal{N}(\hat{\mathbf{x}}_{t|t}, \Sigma^{\mathbf{xx}}_{t|t})$. Given observations at the next time instant $\mathbf{y}_{t+1}$, the classical KF equations update the state distribution to $\mathcal{N}(\hat{\mathbf{x}}_{t+1|t+1}, \Sigma^{\mathbf{xx}}_{t+1|t+1})$ using the following expressions:

$$\begin{aligned} \hat{\mathbf{x}}_{t+1|t+1} &= \hat{\mathbf{x}}_{t+1|t} + \mathbf{K}(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t}) \ , \\ \mathbf{\Sigma}^{\mathbf{xx}}_{t+1|t+1} &= \mathbf{\Sigma}^{\mathbf{xx}}_{t+1|t} - \mathbf{K}\mathbf{\Sigma}^{\mathbf{yy}}_{t+1|t}\mathbf{K}^T \ , \end{aligned}$$

where

$$\mathbf{K} = \mathbf{\Sigma}^{\mathbf{xy}}_{t+1|t} \left(\mathbf{\Sigma}^{\mathbf{yy}}_{t+1|t} \right)^{-1} \ .$$

The computation of this expressions require to characterize:

- the distribution of the system state prediction $\mathcal{N}(\hat{\mathbf{x}}_{t+1|t}, \mathbf{\Sigma}^{\mathbf{xx}}_{t+1|t})$.
- the distribution of the expected observations $\mathcal{N}(\hat{\mathbf{y}}_{t+1|t}, \mathbf{\Sigma}^{\mathbf{yy}}_{t+1|t})$.
- the cross-correlation between the predicted state and observations $\mathbf{\Sigma}^{\mathbf{xy}}_{t+1|t}$.

This distributions are obtained propagating $\mathcal{N}(\hat{\mathbf{x}}_{t|t}, \Sigma^{\mathbf{xx}}_{t|t})$ with equation (4) and (10). When nonlinear system and observation models are concerned, there are no closed-form solutions to estimate them. The UKF approximate them using the Unscented Transform (UT), sketched in figure 8. Given a Gaussian distribution to be transformed, a set of samples are deterministically chosen which match its mean and covariance. This samples are propagated using the nonlinear system, generating a cloud of transformed points. Computing the sample mean and covariance of this points (properly weighted) the transformed Gaussian distribution is characterized. The UKF version used in this paper follows the proposal in [12], which details an efficient implementation of this algorithm based on a generalization of the UT described in [13].

VI. RESULTS

The tracking proposal described in this paper is one of the main modules of a complete framework designed to detect and track vehicles. This framework is composed by:

- a vehicle detector, that analyses frames detecting the presence of new vehicles.
- a target initializer, that given a vehicle detection, estimates its initial 3D pose and dimensions, and adds it to the current tracker state.
- the vehicle tracking module described in this paper.
- a control module, that takes care of suppressing tracked targets which have no longer interest (are occluded, miss-tracked, or out of the camera field of view).

Results obtained from the cooperation of all this modules are encouraging, showing a reliable performance when different types of vehicles are simultaneously tracked under challenging acquisition conditions. However, at this moment the only evaluation that can be provided is qualitative. Work

Fig. 9. Detection-tracking procedure. Frame 000 is the first frame of the sequence. After several consecutive detections (frame 005), a target is added to the tracking module (frame 008). Frame 061 shows that the tracker is robust to sudden changes in the lightening conditions. At frame 110 a new target is added to the tracking module. Frame 174 shows the situation previous to the elimination of one target, due to the occlusion it suffers. Also it is shown how a distant truck is detected. Frame 207 show the simultaneous tracking of two cars and one truck. Frame 260 shows the situation previous to the elimination of one target that leaves the camera field of view.

is in development to evaluate the described proposal using ground truth data. Figure 9 exemplifies in some selected frames the performance of the detection–tracking framework used to test the proposed tracking method.

VII. CONCLUSIONS AND FUTURE WORK

This paper has described a proposal to track the 3D pose and velocity of multiple vehicles, from images acquired from a mobile platform. This problem has been addressed adapting the 3D model in [4] to multiple vehicle tracking. Taking profit of the available a priori knowledge on this particular problem, a novel model of vehicle dynamics has been proposed, which constraint the velocity in which vehicles change their pose, inside a range of physically plausible values. This contributes in a more reliable tracking, as unreasonable vehicle velocities are impossible to be considered. The estimation algorithm used has been the UKF, which supersedes the usually–applied EKF due to its simplicity and bigger accuracy. The proposal has been tested in long sequences, where a reliable performance has been observed even in poor acquisition conditions. However, a more systematic evaluation of the system should be done. It is planned to develop a simulator, in order to generate test sequences with its corresponding ground truth information. This will allow to quantify the precision of the system proposed in controlled scenarios, under different noisy situations. It will also allow to quantify how the performance of the tracker is affected by a wrong initialization of the target pose and dimensions. In a second phase, it is expected to acquire new sequences using a host vehicle equipped with several extra sensors. For each acquired image, it will be registered information about the host movement (provided by a velocity and a steering angle sensor), as well as information of elements on the road ahead (provided by radar and lidar sensors). This information will be compared with the output of the described tracking method, in order to evaluate its performance in real situations.

REFERENCES

[1] E. Dickmanns, "The development of machine vision for road vehicles in the last decade," in *Int. Symp. on Intelligent Vehicles*, Versailles, June 2002.

[2] M. Maurer, R. Behringer, S. Fürst, F. Thomanek, and E. Dickmanns, "A compact vision system for road vehicle guidance," in *Int. Conf. on Pattern Recognition*, vol. 3, August 1996, pp. 313–317.

[3] F. Thomanek and E. Dickmanns, "Obstacle detection,tracking and state estimation for autonomous road vehicle guidance," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 2, Raleigh, 1992, pp. 1399–1406.

[4] F. Dellaert and C. Thorpe, "Robust car tracking using kalman filtering and bayesian templates," in *Proceedings of SPIE: Intelligent Transportation Systems*, vol. 3207, October 1997.

[5] F. Dellaert, D. Pomerleau, and C. Thorpe, "Model-based car tracking integrated with a road-follower," in *IEEE Conf. Robotics and Automation (ICRA)*, 1998, pp. 1889–1894.

[6] M. Betke, E. Haritaoglu, and L. S. Davis, "Real-time multiple vehicle detection and tracking from a movin vehicle," *Machine Vision and Applications*, no. 12, pp. 69–83, 2000.

[7] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[8] J. Bouguet. (2004, October) Camera calibration toolbox for matlab. MRL - Intel Corp. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_ doc/

[9] A. Blake and M. Isard, *Active Contours*. Springer-Verlag, 1998.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[11] S. J. und J. K. Uhlmann, "A new extension of the kalman filter to nonlinear systems," in *The Proceedings of AeroSense: The 11th Int. Symp. on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, Florida, 1997.

[12] R. van der Merwe and E. A. Wan, "The square-root unscented kalman filter for state and parameter-estimation," in *Int. Conf. on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, May 2001.

[13] S. J. Julier, "The Scaled Unscented Transformation," in *Proceedings of the IEEE American Control Conference*. Anchorage AK, USA: IEEE, 8–10 May 2002, pp. 4555–4559.