Vehicle detection and tracking using homography-based plane rectification and particle filtering

Jon Arróspide, Luis Salgado and Marcos Nieto
Grupo de Tratamiento de Imágenes, E. T. S. Ing. Telecomunicación
Universidad Politécnica de Madrid, 28040 Madrid, Spain
{jal, lsa, mnd}@gti.ssr.upm.es

Abstract—This paper presents a full system for vehicle detection and tracking in non-stationary settings based on computer vision. The method proposed for vehicle detection exploits the geometrical relations between the elements in the scene so that moving objects (i.e., vehicles) can be detected by analyzing motion parallax. Namely, the homography of the road plane between successive images is computed. Most remarkably, a novel probabilistic framework based on Kalman filtering is presented for reliable and accurate homography estimation. The estimated homography is used for image alignment, which in turn allows to detect the moving vehicles in the image. Tracking of vehicles is performed on the basis of a multidimensional particle filter, which also manages the exit and entries of objects. The filter involves a mixture likelihood model that allows a better adaptation of the particles to the observed measurements. The system is specially designed for highway environments, where it has been proven to yield excellent results.

I. INTRODUCTION

Nowadays advanced driver assistance systems receive increasing interest both commercially and from the scientific community. In particular, much work has been devoted to the research on techniques for the detection of vehicles in traffic scenarios based on monocular computer vision, due to its low cost and good performance. Aside from traditional knowledge or feature-based approaches, which are dependent on the specific conditions (e.g., weather, illumination), use of motion information is in the basis of many recent works in the field, as this is inherent to the vehicles regardless of the conditions.

Usually movement on a planar surface (i.e., the road) is assumed. Nonetheless, environments using non-stationary camera settings, such as the considered traffic environment, feature motion both of the own vehicle (ego-motion) and of the surrounding vehicles. Many methods have been presented that aim at computing ego-motion in order to decouple these motions [1][2]. However, reliable and efficient ego motion estimation in still and open issue.

As for vehicle tracking, statistical approaches are adopted in many recent works. In particular, particle filters have emerged as a very powerful tool to perform tracking in a wide variety of applications [3][4]. Many of these works assume a given initial detection and perform tracking using likelihood models based on appearance templates. However such detection-by-tracking approaches carry the danger of drifting away from the correct targets [5]. Additionally, tracking is often limited to a constant number of objects.

Some works, such as [6], can handle tracking of a variable number of objects, however they usually require an external detector that manages object entries and exits.

In this work, a complete method for vehicle detection and tracking which addresses the above-mentioned limitations is presented for highway scenarios. The method involves a two-stage procedure. First, the vehicles are detected by using a motion parallax based approach that, as opposed to classical methods, does not require previous ego-motion calculation, which is usually prone to errors. Alternatively, it computes the homography that governs the relations between successive views of the plane [7]. The robustness of the method is ensured by introducing the instantaneous homography measurements into a probabilistic framework based on Kalman filtering. Once a reliable homography is available, the previous image is aligned with the current image, showing the projected position of the vehicles under the assumption that they are static. The difference between aligned images is only significant at those zones of the image featuring motion, that is, the regions where vehicles are. Therefore, the position of the vehicles is obtained from this difference.

On the other hand, tracking of vehicles is achieved through a joint multidimensional state-space particle filter, which is able to handle a variable number of objects. Most remarkably, the management of vehicle entries and exits is realized intrinsically by the designed particle filter, hence avoiding the need for an external control module. The filter achieves thus a two-fold objective: first, it provides a temporal tracking of the measurements provided by the motion-based detector, and second it handles vehicle entries and exits by progressively adapting to the observed measurements on the basis of a complex mixture likelihood model. The model allows to attain smooth tracking of all the vehicles in the scene while removing spurious artifacts produced in the motion segmentation.

II. VEHICLE DETECTION

As stated above, vehicle detection is attained on the basis of road plane alignment between consecutive frames. Namely, alignment of images is performed by warping the previous image by a projective transformation or homography. The steps towards the estimation of this homography

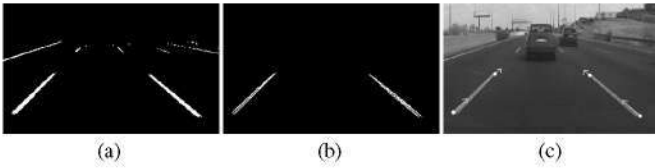<center>(a)   (b)   (c)</center>

Fig. 1. Road feature extraction: (a) response to lane marking detector; (b) segments rendered by Hough Transform applied to (a); (c) road features and their correspondences in the previous image.

and the naturally following image alignment and vehicle detection are explained in the following subsections.

A. Feature extraction and matching

The planar homography between successive time points can be computed from feature correspondences in the road. In this paper we present a novel feature extractor, which is based on the detection of the lane markings in the road. In particular, we propose to restrict the search of feature points to those regions with highest probability to contain lane markings. The presented lane marking detector is an enhancement of that presented in [8], and is specially designed to provide reliable results for the posterior feature matching. Notwithstanding, note that any other feature extraction and matching approach may well be used within the system proposed for vehicle detection and tracking (see [9][10] for other references on road feature extractors). In particular, the homography calculation module in Section II-B works regardless of the used feature extraction method as long as a set of road plane correspondences is provided at its input.

In particular, the proposed lane marking detector is applied to each row of the image, and assumes that the appearance of lane markings in this one-dimensional domain is given by pulses of high intensity values surrounded by darker regions. Therefore, the analysis is done by independently filtering each row j of intensity values, denoted by $\{x_i\}_{i=1}^{W}$, which renders a new filtered data array $\{y_i\}_{i=1}^{W}$, defined as

$$y_i = 2x_i - (x_{i-\tau(j)} + x_{i+\tau(j)}) - |x_{i-\tau(j)} - x_{i+\tau(j)}| \quad (1)$$

where $\tau(j)$ is the width parameter that governs the filtering process and W is the width of the image in pixels. The filter produces high responses for positions with values x_i higher than their neighbors at distance τ to the left and to the right. Note that τ is dependent on the coordinate of the specific row, j: indeed, due to the perspective effect, τ must decrease as j approaches the vanishing point of the lane markings. The last term in (1) penalizes those cases where the difference between left and right neighbors is high (such as vehicle to road transitions or narrow stretches of the road between vehicles or shadows); thus, this filter is less prone to errors than traditional gradient-based detectors.

Fig. 1 (a) shows an example of the thresholded response to the filter. As expected, dense white areas arise in the positions of the actual lane markings. In order to find the underlying lines, a Hough Transform is applied to the filtered image, so

that N clusters of segments are obtained. In the example, it is $N = 2$ (see Fig. 1 (b)). Each cluster of segments corresponds to a lane marking, denoted by M_i ($i = 1, ..., N$), and contains S_i candidate segments. Let $\mathbf{s}_{i,j}$ be a segment j belonging to the lane marking M_i; it is defined as $\mathbf{s}_{i,j} = (\mathbf{p}_{l_{i,j}}, \mathbf{p}_{u_{i,j}})$, where $\mathbf{p}_{l_{i,j}}$ and $\mathbf{p}_{u_{i,j}}$ are its lower and upper end points, respectively. Finally, a representative segment $\mathbf{s}_i$ is extracted for every lane marking as the mean of the segments associated to it: $\mathbf{s}_i = (\mathbf{p}_{l_i}, \mathbf{p}_{u_i}) = \frac{1}{S_i} \sum_{j=1}^{S_i} (\mathbf{p}_{l_{i,j}}, \mathbf{p}_{u_{i,j}})$, for $i = 1, 2, ...N$.

The segments s_i represent the position and orientation of the lane markings on the road. Then, in order to span all over the lane markings, regions are grown centered on these segments, so that the extended regions contain the complete lane markings. These regions constitute the search area for feature points, which are extracted using the Harris corner detector. Generally, features are obtained in the ends of of the lane markings, as shown in the example in Fig. 1 (c). Once a set of reliable features is available, their correspondences on the previous image are obtained through the Lucas-Kanade algorithm. An example of point correspondences is given in Fig. 1 (c) with arrows that point from the current locations of the features to their correspondences in the previous frame.

B. Homography estimation

The homography matrix H defines the transformation of points belonging to the plane from the previous to the current instant as $\mathbf{x}_k = \mathrm{H}\mathbf{x}_{k-1}$, where $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ are the homogeneous coordinates of the features at times $k-1$ and k, respectively. Hence, this homography matrix can be computed using the Direct Linear Transformation (DLT) [7] if sufficient point correspondences between the images are available (i.e., four or more). Nonetheless, the computed homography may be highly affected by the inclusion of bad correspondences in the set of feature points (e.g., the homography is unreliable when only road features near the horizon line are available). In some works, robust estimation techniques are applied to the set of correspondences (i.e., LMS technique in [2]), so that outliers can be identified and filtered. However, this does not ensure the removal of all the outliers. These can be especially harmful when few points are available for the computation of H, as is the case in a traffic environment.

In contrast, the system proposed in this paper is built upon a linear estimation process based on Kalman filtering, which allows to compensate for errors in homography calculation and to reliably estimate the homography at each instant. In effect, the evolution of the parameters of the homography matrix is smooth and the noise observed in the instantaneous measurements is similar-to-Gaussian, as ensured by the Kolmogorov-Smirnov test conducted on test data, hence the use of Kalman filter.

Let us inspect the analytical expression of the homography given the scenario depicted in Fig. 2. The figure considers a vehicle with an on-board camera moving on a flat road plane, $\pi_0 = (\mathbf{n}^\top, d)^\top$, where $\mathbf{n} = (0, 1, 0)^\top$. The world coordinate system has its origin in the position of the camera at instant
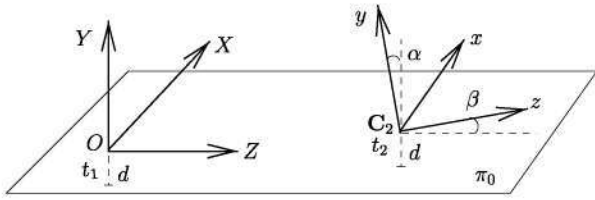
Fig. 2. Relative pose of camera at time points t_1 and t_2 with respect to the world coordinate system (that has its origin at the position of the camera at t_1) and to the plane π_0.

Fig. 3. Image alignment: (a) image at time k of the video sequence; (b) image at time $k-1$ warped with H; (c) difference between aligned images, i.e., (a) and (b).

t_1 and its axes coincide with those of the camera coordinate system. As the vehicle moves forward, at time t_2 rotations might have occurred around the Y-axis, e.g., if the vehicle changes lane or takes a curve, and around the X-axis due to camera shaking. Assuming a pinhole camera model, the camera projection matrices at times t_1 and t_2 are respectively

$$P_1 = K[\mathbf{I}|\mathbf{0}]$$
$$P_2 = KR_x(\alpha)R_y(\beta)[\mathbf{I} - \mathbf{C}_2]$$

where $\mathbf{C}_2$ is the camera position at time t_2. If the speed of the vehicle is v, then $-R_y(\beta)\mathbf{C}_2 = \mathbf{t} = -(0,0,1)^\top v/f_r$, where $f_r = 1/(t_2 - t_1)$ is the frame rate. From the above camera projection matrices, it is straightforward to derive the expression for the homography as [7]

$$H = KR_x(\alpha)(R_y(\beta) - \mathbf{t}\mathbf{n}^\top/d)(KR_c)^{-1}$$

In successive frames, the homography changes only slightly, due to the intrinsic constraints in the motion of the vehicles. Therefore, the evolution of the homography matrix is modeled with a constant-state Kalman filter, whose state vector is composed of the elements H_{ij} of the homography matrix, i.e., $\mathbf{x}_k^\top = \{H_{ij}, 1 \leq i,j \leq 3\}$. Thus, it is $\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{w}_{k-1}$, where $\mathbf{w}_k$ represents the process noise, with Gaussian distribution $p(w) \sim N(0, Q)$. The instantaneous homography matrix, H^k, computed from image correspondences as explained in Section II-A, constitutes the measurement vector as $\mathbf{z}_k^\top = \{H_{ij}^k, 1 \leq i,j \leq 3\}$. It is then $\mathbf{z}_k = \mathbf{x}_k + \mathbf{v}_k$, where $\mathbf{v}_k$ has as well a Gaussian distribution $p(v) \sim N(0, R)$, independent of $\mathbf{w}_k$, representing measurement noise. The measurements are expected to be noisy due to the usually scarce number of points available for solving DLT in traffic environments. Hence, the measurement noise covariance matrix will usually be larger than the process noise covariance (in our system it is $Q = 10^{-6}$, $R = 10^{-3}$). However, note that the intrinsic inaccuracies of the measurement are corrected through the Kalman update stage, that yields a smoothed posterior estimate $\hat{\mathbf{x}}_k$.

On the other hand, the Kalman filter also generates a prediction of the state at time k, which in this case equals the corrected estimate at the previous instant: $\hat{\mathbf{x}}_k^- = \hat{\mathbf{x}}_{k-1}$. This vector contains the expected values of the elements of the homography matrix at time k. Thus, a predicted homography matrix, H^p, may be built with the elements of $\hat{\mathbf{x}}_k^-$. Most importantly, this predicted matrix provides a natural way

to evaluate the goodness of the incoming measurements; indeed, it is natural to accept or dismiss measurements according to their distance to the prediction.

In order to measure the distance between matrices, the 2–norm of a matrix induced by the vectorial norm of a Euclidean space is used. Analogously to the well-known vector norm, this matrix norm, which is given by the largest singular value of the matrix in its Singular Value Decomposition (SVD) [11], provides a measure of the magnitude of the matrix. The distance based on the 2–norm allows us to define a rule for incoming homography acceptance. Specifically, if $\|H^k - H^p\| < t_a$, then the measured homography matrix is accepted and the a priori prediction is updated with the new measurement within the Kalman filter framework. Otherwise, the incoming measurement is not relied on and the system uses the predicted homography for further processing stages. The threshold for homography acceptance, t_a, must be predefined according to kinetic constraints of the vehicles within their environment (see [12]) for an exhaustive discussion on the choice of t_a).

C. Image alignment and vehicle detection

Once a reliable estimate of the homography, H, is obtained as a result of Kalman filtering, image alignment between previous and current frame is performed. This is achieved by warping the previous image with H. An example of image alignment through warping with homography is shown in Fig. 3: in (a) the image at time k of a traffic sequence acquired with an on-board camera is shown; (b) represents the image obtained after warping the image at time $k-1$ with H. The difference between aligned images is expected to be null for all the elements of the road plane except for the contact zones of the vehicles with the road. Fig. 3 (c) illustrates this difference for the previous example. As can be observed, white regions -indicating significant difference- appear in the areas of motion of the vehicles in the road.

The position of moving objects on the plane is extracted by computing the Sum of Absolute Differences (SAD) over the difference between aligned images, and selecting those regions with significant differences. Other measures such as ZCC or ZNCC, which are invariant to brightness and contrast changes, could also be used. However, these changes are negligible between consecutive frames, therefore SAD is preferred due to its lower computational requirements. In particular, images are scanned bottom to up in search of regions with high SAD, which are classified as corresponding to vehicles. Assuming there are M measurements at time k,

the detected set of vehicles is denoted by $\mathbf{z}_k = \{\mathbf{z}_k^j\}_{i=1}^M$, where each measurement $\mathbf{z}_k^j$ is given by the mid-lower point of its associated region of differences.

Note that differences between aligned images might also be produced by background elements above the road plane (e.g., the guardrail in Fig. 3), in which the homography of the road plane does not hold. Therefore, the set of regions classified as vehicles at time k might contain some artifacts, i.e., regions that do not correspond to vehicles. These artifacts usually describe random patterns, therefore they may be removed by analyzing the temporal coherence of the measurements. This issue is addressed in the second stage of the proposed method, namely the tracking stage, which is described in the following section.

III. VEHICLE TRACKING

At this stage we introduce a probabilistic framework, provided by the particle filter, which takes account of the uncertainty in the measurement of the vehicle. Additionally, this framework allows to give coherence to the independent intraframe measurements delivered by the vehicle detector, so that it is possible to track vehicles in time and to analyze their trajectories.

A. Probabilistic tracking framework

The aim of the tracking stage is to update the estimated positions of the vehicles on the receipt of new measurements. The Bayesian approach is ideally suited for such dynamic state estimation problem, as it allows to update the state recursively based on a probabilistic formulation. In particular, particle filters aim at approximating the posterior probability density function of the state with a set of random *particles*, and their associated importance weights. The posterior density at time k can be approximated as [13]

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{s=1}^{n_s} w_k^{(s)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(s)})$$

where $w_k^{(s)}$ is the weight associated to sample $\mathbf{x}_k^{(s)}$. Particularly, in this work the SIR (Sampling Importance Resampling) particle filter is used, which involves a resampling step at every time index as explained in [13].

In the proposed framework the measurement vector is given by the observed positions of the vehicles, $\mathbf{z}_k$, delivered by the detection module explained in Section II. As for the system model, in contrast to usual random walk assumption, it is designed to be linear with constant velocity. This is in line with the locally smooth motion of vehicles in the road. Hence, the state vector of the system is composed of the position, $\mathbf{x}_k^i = \{x_k^i, y_k^i\}$, and velocity, $\dot{\mathbf{x}}_k^i = \{\dot{x}_k^i, \dot{y}_k^i\}$ of each one of the N tracked vehicles, i.e., $\mathbf{x}_k = \{x_k^i, y_k^i, \dot{x}_k^i, \dot{y}_k^i\}_{i=1}^N$. On the other hand, the measurement model is defined as

$$p(\mathbf{z}_k|\mathbf{x}_k) = \prod_{j=1}^M p(\mathbf{z}_k^j|\mathbf{x}_k) = \prod_{j=1}^M \sum_{i=1}^N [\alpha_{ij} p_i(\mathbf{z}_k^j|\mathbf{x}_k^i) + \alpha_u \mathcal{U}(\mathbf{z}_k^j)] \quad (2)$$

Let us inspect this expression. First, the measurements are considered to be independent from one another, thus the joint probability density is the product of the individual densities.

Besides, the probability density of each measurement is modeled by a mixture model [14]. In (2), the mixture model mixes distributions, p_i, with mixing coefficients α_{ij}. These are modeled as bivariate Gaussian distributions centered around $\mathbf{x}_k^i$. Indeed, each measurement might correspond to any of the existing vehicles, hence the use of a mixture model, in which each model, p_i, gathers the probability that the measurement was generated by the i-th vehicle. However, note that the model in (2) includes an additional term given by a uniform distribution, $\mathcal{U}(\cdot)$. This term is added (with mixing coefficient α_u) in order to model scattered noise, that is, to represent the probability that the measurement is actually produced by noise rather than by any of the vehicles.

The mixing coefficients of the mixture model, α_{ij}, may be chosen arbitrarily as long as $\sum_{i=1}^N \alpha_{ij} + \alpha_u = 1 \; \forall j$. A natural decision is to make them proportional to the likelihood $p_i(\mathbf{z}_k^j|\mathbf{x}_k^i)$ of the vehicle i to which they are associated:

$$\alpha_{ij} = (1 - \alpha_u)p_i(\mathbf{z}_k^j|\mathbf{x}_k^i)/\sum_{t=1}^N p_i(\mathbf{z}_k^j|\mathbf{x}_k^t) \quad \left\{ \begin{array}{l} i = 1, \dots, N \\ j = 1, \dots, M \end{array} \right.$$

The designed mixture model allows that, given one measurement, the likelihood of the state vector sums the probability that each one of the vehicles in the state vector generated the measurement, rather than only considering the likelihood of the vehicle closer to the measurement, as done in typical rule-based approaches. In addition, the coefficient α_u weighs the importance of the scattered noise component, which will be tuned according to the expected measurement process noise, as explained in Section III-B. Both aspects maximize the flexibility of the observation model and its adaptivity to the measurements compared to classical methods.

Given the motion and measurement models defined above, the operation of the SIR particle filter for every time step is summarized as follows. First, the particles from previous time point are propagated according to the system model. Then the weight of each particle is computed. In SIR, these weights are given by the likelihood in (2). Particles are resampled according to the procedure described in [13]. Finally, at each instant the vehicle position estimates are inferred as the mean of the particles, i.e., $\bar{\mathbf{x}}_k = \frac{1}{n_s} \sum_{s=1}^{n_s} \mathbf{x}_k^{(s)}$.

B. Management of vehicle entries and exits

In this work we propose to exploit the strength of the particle filter not only as a means for object tracking but also to manage the entry and exit of vehicles. For vehicle entries, the underlying idea is to hypothesize an extension of the state vector (that is, to hypothesize a new vehicle) when a new measurement arises. Specifically, suppose that at time k we have a measurement vector $\mathbf{z}_k$. If any of the elements $\mathbf{z}_k^j$ of the vector is not adequately represented by the current set of vehicles in the state vector, a transitory period is triggered. At the start of this period, a small subset of n_r particles from the total pool of n_s particles will be devoted to hypothesize the existence of a new vehicle. Namely, n_r arbitrary particles (recall that after the resampling step all particles have the same weight) are eliminated and substituted by new particles

with a larger dimension, $N+1$. The state vector of these particles, denoted by a prime, comprises the mean estimate of the existing vehicles and a new element, $\mathbf{x}_k^{N+1}$, given by the new measurement $\mathbf{z}_k^j$:

$$\mathbf{x}'_k = \{\bar{\mathbf{x}}_k, \mathbf{x}_k^{N+1}\}$$

During the transitory state, both N-dimension and $(N+1)$-dimension particles are propagated with the motion model. The weights of the particles are then evaluated according to the measurement model in (2). Naturally, the particles that better match the observed set of vehicles will have larger weights and thus dominate the resampling step. In particular, if incoming observations include a measurement for the hypothesized new vehicle, the $(N+1)$-dimension particles (or simply the $(N+1)$-particles) will have larger weights, therefore they will propagate faster and eventually outnumber the N-particles. In contrast, if the hypothesized vehicle is not supported by incoming measurements, N-particles will propagate over $N+1$-particles, which will eventually fade away. The transitory period stops when the number of particles of either N-dimension or $(N+1)$-dimension falls behind a predefined proportion. At this point, all particles take the same dimension and the filter returns to stationary state.

The interpretation of the particle filter behavior regarding new vehicle management is straightforward. Namely, the propagation of $(N+1)$-particles is a consequence of the observation of persistent and time-coherent measurements in a specific region of the image, hence the eventual confirmation of a new vehicle. Conversely, the non-propagation of these particles implies spurious or non-coherent measurements, which are effectively ignored by the particle filter, as requested, hence preventing the system from producing false detections.

With regards to vehicle exits, these are inherently managed by the proposed tracker. Namely, a vehicle is tracked while its position inferred by the particle filter is within the region of interest, and it is removed from the tracked vehicle set whenever it abandons this region.

As a final remark, note that the selection of the coefficient α_u in the mixture model in (2) is tightly related to the dimensionality of the state vector. This coefficient models the probability that measurements are produced by noise rather than by actual vehicles. In effect, the transitory state is triggered by the observation of measurements that are not well represented by the posterior pdf. These could be produced by new vehicles, or by artifacts in the measurement generation process. Hence, it is natural to account for this degree of uncertainty in the measurements by increasing the coefficient α_u. As a consequence, during the transitory state particles will slowly adapt to measurements, so that only consistent and persistent measurements will lead to the detection of new vehicles. In contrast, during stationary state, the measurements are in line with the posterior pdf, thus making the measurements process reliable and implying the reduction of the coefficient α_u.
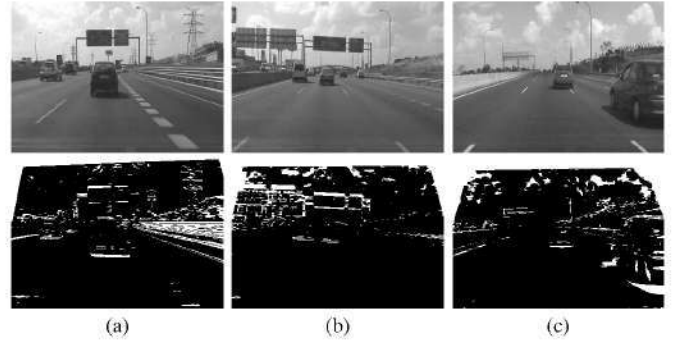

Fig. 4. Results of the proposed motion-based segmentation for three different scenarios. The corresponding input image is provided in the upper row as a reference for each example. Note that differences (white) in the road plane correspond mainly to the regions of motion of the vehicles.

IV. EXPERIMENTS AND DISCUSSION

The proposed system has been tested for video sequences involving a wide variety of driving situations, in which it is proved to yield excellent results. Overall, an average detection rate of above 90% is obtained for a set of scenarios, including different illumination, weather and traffic conditions (except for severe rain and night-time, in which the rate decreases). The system operates at a frame rate of between 5 and 10 fps.

Fig. 4 gathers some example results of the proposed motion-based segmentation. Specifically, it comprises the original image (upper row) and the obtained difference between aligned images (lower row) for three different scenarios. In all cases significant differences are observed in the regions of motion of the vehicles within the road plane (naturally, due to the perspective effect the region of differences is wider for near vehicles). Indeed, compared to classical methods for detecting vehicles, such as those based on vehicle structure (e.g., edges, symmetry, etc.) or application specific features (e.g., analysis of vehicle shadows) the proposed method for detecting vehicles is more invariant to the particular scene conditions (e.g., weather, time of the day) since it is based on a feature inherent to the vehicles, i.e., motion. For instance, a method based on shadow detection would only detect vehicles overtaking the own vehicle once these vehicles are completely within the limits of the image, while an almost immediate detection is achieved with the proposed method (see Fig. 4 (c)).

However, note the existence of artifacts caused by background elements. This is effectively solved by the designed tracking framework based on particle filtering. Consider the image sequence shown in the upper row of Fig. 5. A segmentation artifact in the right-hand side of the image at time $k-2$ generates a set of particles containing an object in this zone (highlighted by a dashed rectangle). This artifact does not appear at times $k-1$ and k, therefore the likelihood model of the particle filter penalizes the above-mentioned particles, which decrease at time $k-1$ and finally vanish. Analogously, in the example in the lower row of Fig. 5, the measurement at time $k-2$ triggers new particles. However,
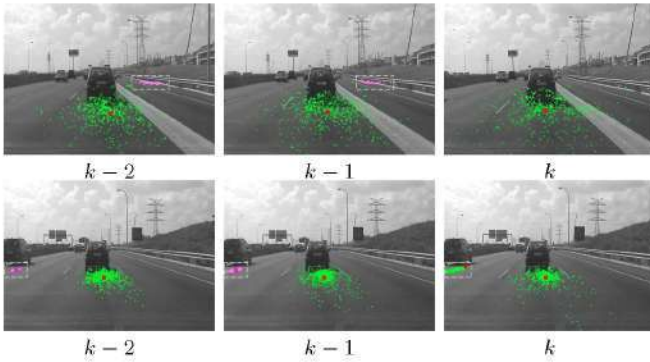
Fig. 5. Management of object entries performed by the particle filter. The state of each particle is represented by green dots. New hypothesized vehicles during transitory period are painted in pink and highlighted with a dashed square. The red mark indicates the position estimate for each vehicle.
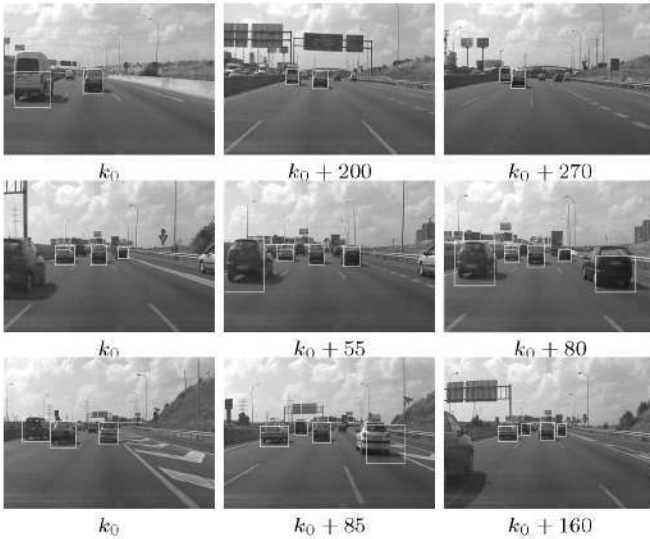

Fig. 6. Tracking for three example sequences is shown in the upper, middle and lower row, respectively. Observe that in all cases appearing vehicles are swiftly detected and and existing vehicle are correctly tracked.

V. CONCLUSIONS

In this paper a full system for vehicle detection and tracking in highway environments given by an on-board camera has been presented. The system first detects vehicles on the basis of successive image rectification using plane-to-plane homographies. The framework presented for homography estimation and image alignment has been proven to be robust to errors in the measurement process, thus providing reliable and accurate vehicle detection. On the other hand, a method for tracking of instantaneous vehicle detections has been presented on the basis of an adaptive multidimensional particle filter. This filter is able not only to provide robust tracking of the vehicles in the image, but also to efficiently handle entry and exit management. The combined detection and tracking strategy yields excellent results while circumventing the limitations of state of the art methods.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] K. Yamaguchi, T. Kato and N. Ninomiya, "Moving Obstacle Detection using Monocular Vision", *in IEEE Intelligent Vehicles Symposium*, Tokyo, Japan, 2006, pp. 288–293.

[2] H. Zhou, A. M. Wallace and P. R. Green, A Multistage Filtering Technique to Detect Hazards on the Ground Plane, *Pattern Recogn. Lett.*, vol. 24, no. 9, 2003, pp. 1453–1461.

[3] J. J. Pantrigo, A. Sánchez, A. S. Montemayor and A. Duarte, Multi-dimensional Visual Tracking using Scatter Search Particle Filter, *Pattern Recognition Letters*, vol. 29, no. 8, 2008, pp. 1160–1174.

[4] T. Gao et al., Moving Vehicle Tracking based on SIFT Active Particle Choosing, *Lecture Notes in Computer Science, Advances in Neuro-Information Processing*, vol. 5507, 2009, pp. 695–702.

[5] A. Ess, B. Leibe, K. Schindler and L. van Gool, Robust Multiperson Tracking from a Mobile Platform, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, 2009, pp. 1831–1846.

[6] Z. Zivkovic, A. T. Cemgil and B. Kröse, Approximate Bayesian Methods for Kernel-based Object Tracking, *Computer Vision and Image Understanding*, vol. 113, no. 6, 2009, pp. 743–749.

[7] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK; 2000.

[8] M. Nieto et al., "Stabilization of Inverse Perspective Mapping Images based on Robust Vanishing Point Estimation", *in IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, 2007, pp. 315–320.

[9] T. Veit, J.-P. Tarel, P. Nicolle and P. Charbonnier, "Evaluation of Road Marking Feature Extraction", *in IEEE Conf. on Intelligent Transportation Systems*, Beijing, China, 2008, pp. 174–181.

[10] C. Boussard et al., "Vehicle Dynamics Estimation for Camera-based Visibility Distance Estimation", *in IEEE/RSJ Conf. on Intelligent Robots and Systems*, Nice, France, 2008, pp. 600–605.

[11] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, Englewood Cliffs, NJ; 1999.

[12] J. Arróspide, L. Salgado, M. Nieto and R. Mohedano, Homography-based Ground Plane Detection Using a Single On-board Camera, *IET Intelligent Transport Systems*, vol. 4, no. 2, 2010 (in press).

[13] S Arulampalam et al., A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking, *IEEE Trans. on Signal Processing*, vol. 50, no. 2, 2002, pp. 174–188.

[14] C. M. Bishop, *Pattern Recognition and Machine Learning, Ch. 9: Mixture Models and EM*, Springer Science+Business Media, LLC, New York, NY; 2006.

this measurement is generated by a vehicle entering in the left-hand side, thus it produces coherent measurements in the following time indexes. Consequently, particles comprising a track for this vehicle grow and finally the vehicle is confirmed.

The positions of the vehicles are estimated as the mean of the particles. Results of the method are shown in Fig. 6 for three different scenarios, proving its ability to reliably track multiple vehicles. Most remarkably, the predictive nature of the homography estimation process enables the system to operate for long stretches of time without new incoming measurement. Indeed, the homography predicted by the Kalman filter is available at every time step and can be used to attain image alignment and posterior vehicle detection. This is the case for the example in the lower row of Fig. 6, in which the predicted homography is used due to the lack of new measurements; however, correct tracking of the vehicles is kept with a small impact on the accuracy of the detections.