Homography-based ground plane detection using a single on-board camera

J. Arróspide L. Salgado M. Nieto R. Mohedano

Grupo de Tratamiento de Imágenes – E.T.S. Ing. Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
E-mail: jal@gti.ssr.upm.es

Abstract: This study presents a robust method for ground plane detection in vision-based systems with a non-stationary camera. The proposed method is based on the reliable estimation of the homography between ground planes in successive images. This homography is computed using a feature matching approach, which in contrast to classical approaches to on-board motion estimation does not require explicit ego-motion calculation. As opposed to it, a novel homography calculation method based on a linear estimation framework is presented. This framework provides predictions of the ground plane transformation matrix that are dynamically updated with new measurements. The method is specially suited for challenging environments, in particular traffic scenarios, in which the information is scarce and the homography computed from the images is usually inaccurate or erroneous. The proposed estimation framework is able to remove erroneous measurements and to correct those that are inaccurate, hence producing a reliable homography estimate at each instant. It is based on the evaluation of the difference between the predicted and the observed transformations, measured according to the spectral norm of the associated matrix of differences. Moreover, an example is provided on how to use the information extracted from ground plane estimation to achieve object detection and tracking. The method has been successfully demonstrated for the detection of moving vehicles in traffic environments.

1 Introduction

Computer vision-based analysis has been a central research line to attain scene characterisation for the last decades. Particularly, many works have been devoted to solve problems such as obstacle detection, road detection, ego-motion estimation, localisation etc. when the observer is moving. In general, the moving platform and the objects in the scene are assumed to be moving along a planar surface (e.g. the road). The detection of the region of the image corresponding to this surface is a basic step towards the resolution of the aforementioned problems, especially object detection. Therefore several works have been proposed that aim at estimating the ground plane as the basis for a posterior object detection on that plane [1–3]. Usually, they use a calibrated stereo rig that allows for a direct estimation of the ground region [4, 5]. Nonetheless, stereo systems have a number of disadvantages compared to monocular systems, especially in terms of cost and flexibility.

As for monocular systems, methods accounted in the literature require a complete estimation of the 3D ground plane parameters as well as the 3D camera motion parameters (ego-motion) [1]. Then they compute the homography of the ground plane between two consecutive images based on the expression of the homography between two planes. In this context, the use of inertial sensors (e.g. odometers, speedometers, accelerometers, gyroscopes) has been used in numerous works to facilitate ego-motion estimation [6, 7]. However, this information may not be available in all applications, and even if available it might be inaccurate because of drifts [5, 8], thus it is interesting and challenging to capitalise on all the information attainable via pure vision analysis. Unfortunately, visual ego-motion estimation is a complex task and still an active research line [8–10].

In this paper, a purely vision-based method using a single on-board camera is proposed for the estimation of the ground plane. The proposed method does neither require prior

ego-motion estimation nor explicit extraction of the 3D ground plane parameters. Alternatively, it is based on a robust computation of the homography of the ground plane between two consecutive images from reliable ground plane point correspondences. As opposed to the methods accounted in the literature, which estimate the homography independently for every image frame, the proposed method takes advantage of the temporal coherence of the interframe plane-to-plane homography to construct a probabilistic prediction framework based on Kalman filtering for the computation of the homography.

In the literature there are plenty of works that use Kalman theory for background estimation and foreground detection [11, 12]. These works model the background dynamics of each pixel with a Kalman filter in order to continuously adapt to variations in the background, for example because of illumination changes. Extended Kalman filtres have also been used for vehicle modelling in traffic surveillance scenarios [13]. Here we propose to extend the use of Kalman theory to non-stationary settings. In this case, the Kalman filter will be used to smoothly adapt to changes in the plane-to-plane homography so that a reliable estimation of the ground plane can be attained at every time point.

Ground plane detection is the basis for the posterior detection of objects moving on that plane. In this work, an example is given on how the ground plane information can be used to detect and track objects. This example illustrates the potential of the method proposed for ground plane estimation and its applicability to object detection. The method may be used in any environment featuring a ground plane, although this paper focuses on its application to traffic environments.

2 Overview

The basic idea behind the proposed approach is that pixels that belong to the ground plane have coherent motion patterns when the acquisition platform (e.g. the vehicle) is moving. On the other hand, when moving objects appear on the ground plane, those pixels of the objects that belong to this plane will present a non-coherent motion compared with the rest of the pixels in the plane. The motion of the ground plane points is to be characterised by the planar homography between two consecutive views of the scene. This homography is mathematically expressed by [14]

$$\boldsymbol{H} = \boldsymbol{K}(\boldsymbol{R} + \boldsymbol{R}\boldsymbol{C}\boldsymbol{n}^{\top}/d)\boldsymbol{K}^{-1} \tag{1}$$

where $\boldsymbol{K}$ is the camera calibration matrix, $\boldsymbol{n}$ is the normal vector of the ground plane, $\boldsymbol{R}$ and $\boldsymbol{C}$ are the relative rotation and translation between views and d is the distance between the camera and the ground plane. Most homography-based approaches found in the literature aim to compute the parameters in (1) so as to obtain an analytical expression of the homography for every instant. However, this involves solving two complex problems: ego-motion estimation and 3D ground plane extraction. Ego motion is usually derived from feature correspondences between images [2, 3]. However, in order to compute the ego-motion only features corresponding to static objects should be taken into account. Hence, a method has to be defined first to filter out features belonging to moving objects [2]. On the other hand, and regarding particularly the traffic scenario, roads tend to contain very few feature points, while many points appear in the background objects (buildings, trees etc.) and moving objects (which in turn aggravates the first problem). The non-homogeneous distribution of feature points jeopardises the accuracy of the motion estimation.

Alternatively, in this work a new method for homography calculation is proposed to address the aforementioned problems. First, the method calculates the planar homography directly from feature correspondences rather than previously computing ego-motion. Then, in contrast to existing methods, which deliver an independent calculation of the homography for every instant, the method presented herein involves a linear data estimation framework that provides a time-filtered estimation of the homography. In addition, an outlier rejection technique is built upon this estimation framework, which removes erroneous measurements based on the computation of the spectral norm of the matrix of differences. Finally, alignment of successive images is achieved using this homography and thus the ground region can be detected.

3 Ground plane estimation

As explained above, homography calculation for every pair of consecutive frames is based on feature correspondences. The first requisite is therefore to find a set of reliable feature points lying on the ground plane. Usually, corner detectors (e.g. Harris, KLT) are utilised to extract features, followed by a robust estimation technique (i.e. RANSAC) in which the dominant homography is estimated [2]. However, this approach is not suitable for some environments (e.g. traffic scenarios), since such techniques would render few points on the homogeneous road compared to moving objects and other objects on the background, which have richer corner contents. Consequently, the ground plane homography would probably not be dominant. Here, we propose to use some a priori knowledge of the scenario so that it is possible to obtain feature points that belong to the ground plane.

In effect, the observation of feature points in the ground plane depends on the nature of the specific environment, and therefore the procedure to obtain ground plane feature points must be designed according to it. For instance, in this work, a priori knowledge lies on the existence of lane markings painted on the road. Hence, a lane marking detector as in [15] is used to first localise the regions containing lane markings, and then feature correspondences are sought within these regions. In contrast, in indoors

robot navigation applications the texture of the surface will probably render these correspondences. This can be easily extended to other man-made environments, which typically contain sets of ortho-parallel lines whose structure may be used as prior information. Even for relatively homogeneous surfaces, techniques exist that are able to produce correspondences between images, such as SIFT [16]. In many cases the extracted feature correspondences will be few or imprecise. However, the robustness of the prediction framework compensates for the inherent inaccuracy of the calculated instantaneous homography.

3.1 Homography estimation framework

The transformation of points on the ground plane between images at times $k-1$ and k is given by a planar homography (see (1)) as follows

$$\boldsymbol{x}_k = \boldsymbol{H}\boldsymbol{x}_{k-1} \tag{2}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k-1}$ are the homogeneous coordinates of the features in the current and the previous image, respectively. The planar homography $\boldsymbol{H}$ in (1) consists of eight independent coefficients. Hence at least eight equations (i.e. four-point correspondences) are needed to solve the linear system [14]. The homography may thus be computed if four or more point correspondences are found between images by the means referred in the previous section. Nevertheless, an instantaneous homography computed this way may be corrupted because of inaccurate or erroneous correspondences. This is especially harmful when few points are utilised in the computation of $\boldsymbol{H}$, as is the case in a traffic environment, on account of the scarce number of feature points on the road.

In this work, the estimation of the homography is modelled as a linear process and controlled by means of a Kalman filter [17]. In effect, the homography between ground planes from frame to frame depends only on the position of the camera relative to the ground plane and the displacement and rotation of the camera. Those should vary slowly and thus the difference in $\boldsymbol{H}$ between successive frames shall be very small.

Specifically, let us consider a vehicle moving on a flat road plane π_0, as shown in Fig. 1. The road plane has coordinates $\pi_0 = (\boldsymbol{n}^\top, d)^\top$, where $\boldsymbol{n} = (0, 1, 0)^\top$. The camera looks forward to the road with a small initial rotation, $\boldsymbol{R}_c$, with respect to the world coordinate system. If we consider a vehicle heading forward at time t_1, then at time t_2 a rotation $\boldsymbol{R}_y(\beta)$ may have occurred around the Y-axis with a yaw angle β, for example, if the vehicle changes lane or takes a curve. Additionally, a rotation $\boldsymbol{R}_x(\alpha)$ around the X-axis models changes in the pitch angle α because of possible car bumping. The roll angle is assumed to be zero. Then, the camera projection matrices at times t_1 and t_2 are

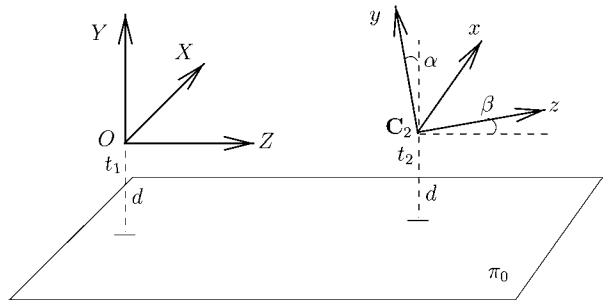

Figure 1 *Relative pose of camera at time points t_1 and t_2 with respect to the world coordinate system (that has its origin at the position of the camera at t_1) and to the plane π_0. In the figure R_c is assumed to be the identity matrix for simplicity*

respectively

$$\boldsymbol{P}_1 = \boldsymbol{K}\boldsymbol{R}_c[\boldsymbol{I}|\boldsymbol{0}]$$
$$\boldsymbol{P}_2 = \boldsymbol{K}\boldsymbol{R}_c\boldsymbol{R}_x(\alpha)\boldsymbol{R}_y(\beta)[\boldsymbol{I}| - \boldsymbol{C}_2]$$

where $\boldsymbol{C}_2$ is the camera position at time t_2. If the speed of the vehicle is v, then $-\boldsymbol{R}_y(\beta)\boldsymbol{C}_2 = \boldsymbol{t} = -(0, 0, 1)^\top v/f_r$, where $f_r = 1/(t_2 - t_1)$ is the frame rate. Given $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$, the homography matrix between planes is the following [14]

$$\boldsymbol{H} = \boldsymbol{K}\boldsymbol{R}_c\boldsymbol{R}_x(\alpha)(\boldsymbol{R}_y(\beta) - \boldsymbol{t}\boldsymbol{n}^\top/d)\boldsymbol{R}_c^{-1}\boldsymbol{K}^{-1} \tag{3}$$

Note that the homography $\boldsymbol{H}$ depends on the camera calibration matrix $\boldsymbol{K}$. This contains the internal camera parameters and is thus specific for each camera model. In order to detach the analysis from the specific camera and to build a general rule for the algorithm parameters, here we define a normalised homography, $\bar{\boldsymbol{H}}$, which removes the dependency on $\boldsymbol{K}$, as

$$\bar{\boldsymbol{H}} = \boldsymbol{K}^{-1}\boldsymbol{H}\boldsymbol{K} = \boldsymbol{R}_c\boldsymbol{R}_x(\alpha)(\boldsymbol{R}_y(\beta) - \boldsymbol{t}\boldsymbol{n}^\top/d)\boldsymbol{R}_c^{-1} \tag{4}$$

In this work we will refer the main definitions in the analysis framework to the normalised homography $\bar{\boldsymbol{H}}$ to achieve the maximum degree of generality, although analogous derivations can also be done using $\boldsymbol{H}$ and the known $\boldsymbol{K}$ of the specific camera. Typically, the camera rotations are small, thus $\boldsymbol{R}_x \simeq \boldsymbol{R}_y(\beta) \simeq \boldsymbol{I}$ and the normalised homography is close to

$$\bar{\boldsymbol{H}} = \boldsymbol{R}_c(\boldsymbol{I} - \boldsymbol{t}\boldsymbol{n}^\top/d)\boldsymbol{R}_c^{-1} \tag{5}$$

As will be shown in Section 3.2, even when rotations occur, the variation in the homography between successive frames is small. In addition, a Kolmogorov–Smirnov test [18] with a typical significance level of 5% has been performed over the elements of the homography matrix in order to evaluate the adequacy of the data to a Gaussian distribution. The test supports the Gaussianity hypothesis. Hence, $\bar{\boldsymbol{H}}$ is introduced into a Kalman filter in which the state vector $\boldsymbol{x}_k$ is composed of the rows $\boldsymbol{h}_j$ of $\bar{\boldsymbol{H}}$. The static process is thus

given by a position-only model with an identity transition matrix

$$\begin{aligned} \boldsymbol{x}_k &= (\boldsymbol{h}_1 \boldsymbol{h}_2 \boldsymbol{h}_3)^\top \\ \boldsymbol{A} &= \boldsymbol{I}_{9\times 9} \\ \boldsymbol{x}_k &= \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1} \end{aligned} \tag{6}$$

where $\boldsymbol{w}_k$ is a Gaussian distribution modelling the process noise. In turn, the measurement vector takes the values of the homography matrix computed from the linear equation system given by the set of corner correspondences in (2). This must be normalised as in (4). Let us denote the instantaneous homography matrix from (2) as $\boldsymbol{H}^i$, and the normalised instantaneous matrix as $\bar{\boldsymbol{H}}^i$. The measurement vector is composed of the rows $\boldsymbol{h}_j^i$ of $\bar{\boldsymbol{H}}^i$

$$\begin{aligned} \boldsymbol{z}_k &= (\boldsymbol{h}_1^i \boldsymbol{h}_2^i \boldsymbol{h}_3^i)^\top \\ \boldsymbol{B} &= \boldsymbol{I}_{9\times 9} \\ \boldsymbol{z}_k &= \boldsymbol{B}\boldsymbol{x}_k + \boldsymbol{v}_k \end{aligned} \tag{7}$$

where $\boldsymbol{v}_k$ is a Gaussian distribution, independent of $\boldsymbol{w}_k$, modelling the measurement noise. Note that some of the instantaneous measurement matrices may be incorrect because of inconsistent correspondences. However, the linear filtering method used renders a prediction of the state vector, that is, the elements of the homography. Following the notation in [17] the prediction of the state vector $\boldsymbol{x}_k$, denoted by $\hat{\boldsymbol{x}}_k^-$, is obtained from the estimate of the state vector in the previous instant, $\hat{\boldsymbol{x}}_{k-1}$, as

$$\hat{\boldsymbol{x}}_k^- = \boldsymbol{A}\hat{\boldsymbol{x}}_{k-1} \tag{8}$$

The predicted state vector contains the expected values of the elements of the homography matrix (hereafter the term homography will be referred to the normalised homography, unless otherwise stated). Analogously to (6), and using a notation coherent with [17], let $\hat{\boldsymbol{x}}_k^-$ be rewritten as $\hat{\boldsymbol{x}}_k^- = (\boldsymbol{h}_1^- \boldsymbol{h}_2^- \boldsymbol{h}_3^-)$. We define a predicted homography matrix, $\bar{\boldsymbol{H}}^{\mathrm{p}}$, built up with the elements of $\hat{\boldsymbol{x}}_k^-$

$$\bar{\boldsymbol{H}}^{\mathrm{p}} = \begin{pmatrix} \boldsymbol{h}_1^- \\ \boldsymbol{h}_2^- \\ \boldsymbol{h}_3^- \end{pmatrix} \tag{9}$$

3.2 Homography update rule

Clearly, a rule may be constructed to compare the instantaneous measurement of the homography with the predicted homography. Accordingly, homographies that significantly differ from the expected transformation can be removed. For this purpose we make use of a norm of a matrix induced by the vectorial norm of a Euclidean space, denoted two-norm or spectral norm. This norm is a natural extension of the concept of norm for a vector, and gives a measure of the magnitude of a matrix. Namely, the two-norm of a matrix $\boldsymbol{A}$ is given by its largest singular value or equivalently by [19]

$$\|\boldsymbol{A}\|_2 = \sqrt{\lambda_{\max}(\boldsymbol{A}^\dagger \boldsymbol{A})} \tag{10}$$

where $\lambda_{\max}$ is the biggest eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{A}^\dagger$ denotes the conjugate transpose of $\boldsymbol{A}$. We apply the induced two-norm to the matrix difference between the measured and the predicted homographies. Only those homographies whose difference to the prediction has a norm below a predefined threshold are accepted

$$\text{if}(\|\bar{\boldsymbol{H}}^{\mathrm{i}} - \bar{\boldsymbol{H}}^{\mathrm{p}}\| < t_n) \Longrightarrow \text{update } \bar{\boldsymbol{H}} \tag{11}$$

The threshold t_n depends on the change of the velocity and the rotation parameters and must therefore be estimated according to the application. In particular, in this work the maximum expectable difference between homography matrices at times k_1 and k_2 has been analysed for the road environment to set the threshold t_n. Let us assume a velocity v_1 at time k_1 and no rotation between times $k_1 - 1$ and k_1. Then, the planar homography between times $k_1 - 1$ and k_1 is $\bar{\boldsymbol{H}}_1 = \boldsymbol{R}_{\mathrm{c}}(\boldsymbol{I} - \boldsymbol{t}_1 \boldsymbol{n}^\top / d)\boldsymbol{R}_{\mathrm{c}}^{-1}$ as in (5), where $\boldsymbol{t}_1 = -(0, 0, 1)^\top v_1 / f_r$. In turn, let us consider that the rotations at time k_2 with respect to $k_2 - 1$ are modelled by $\boldsymbol{R}_y(\beta)$ and $\boldsymbol{R}_x(\alpha)$, and the velocity is $v_2 = v_1 + \Delta v$. Hence, the homography $\bar{\boldsymbol{H}}_2$ between planes at times $k_2 - 1$ and k_2 is given by $\bar{\boldsymbol{H}}_2 = \boldsymbol{R}_{\mathrm{c}}\boldsymbol{R}_x(\alpha)(\boldsymbol{R}_y(\beta) - \boldsymbol{t}_2 \boldsymbol{n}^\top / d)\boldsymbol{R}_{\mathrm{c}}^{-1}$, where $\boldsymbol{t}_2 = -(0, 0, 1)^\top (v_1 + \Delta v)/f_r$.

The differences in the homographies are produced by changes in the velocity and rotation parameters. Regarding velocity, most nation governments enforce a maximum speed limit of $v = 120\,\mathrm{km/h}$ (33.3 m/s). As for the rotation parameters, a maximum rotation of $\alpha = \pm 5^\circ$ around the X-axis because of bumping will be considered. The maximum rotation of a vehicle around the Y-axis will occur in curves. To find the upper bound for this rotation angle, let us consider a circular model of the curve with radius r. The difference in the orientation angle of the vehicle between time points k and $k - 1$, denoted by β, is given by the tangents to the curve at the positions of the vehicles. Fig. 2 synthesises an aerial view of a vehicle taking a curve to the left. The vehicle moves from point A at time $k - 1$ to point B at time k, describing an arc of length s. According to standard road geometry design rules [20], the

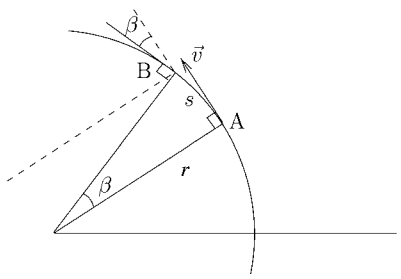

Figure 2 *Derivation of the maximum vehicle rotation in a curve between successive time points*

minimum radius of curvature for motorways at a speed of 120 km/h, assuming a side friction factor of 0.09 and a superelevation of 6%, is $r_{\min} = 875$ m. The objective is to find the upper bound for $\beta = s/r$ in a normal driving situation. As the rotation angle between two consecutive frames will be very small, let s be approximated to $s \simeq v \cdot \Delta t$. The time difference between frames, Δt, depends on the frame rate, and must be well below 1 s (no correspondence between frames will be found beyond this time difference). Hence

$$\beta < \frac{v}{r_{\min}} = \frac{33.3}{875} = 0.038\,\text{rad} = 2.18^\circ \tag{12}$$

In Fig. 3, $\|\bar{\boldsymbol{H}}_1 - \bar{\boldsymbol{H}}_2\|$ is analysed as a function of these parameters between the bounds derived above. First, the norm of the difference is plotted in Fig. 3*a* changing the velocity parameter and assuming zero rotation. In Fig. 3*b*, the difference is analysed for different rotation angles around the *Y*-axis, that is, rotations produced by left or right turns of the vehicle, at a standard speed of 100 km/h. Analogously, the effect of rotations in *X*-axis, which are due to camera bumping, is reflected in Fig. 3*c*. The joint effect of rotation in *Y*- and *X*-axis is shown in Fig. 3*d*. As can be observed, the largest difference occurs with $\alpha = 5^\circ$, $|\beta| = 3^\circ$. In Fig. 3*e*, the same figure as in (a) is plotted with $\alpha = 5^\circ$, $|\beta| = 3^\circ$. Finally, in Fig. 3*f*, $\|\bar{\boldsymbol{H}}_1 - \bar{\boldsymbol{H}}_2\|$ is evaluated for different rotation angles setting the velocity parameter to the lowest value within the expected range, that is, 60 km/h.

Note that for every combination of α, β and Δv in the graphics, it is always $\|\bar{\boldsymbol{H}}_1 - \bar{\boldsymbol{H}}_2\| < 0.1$. Hence, the threshold t_n in (11) is set to $t_n = 0.1$ for the traffic scenario in highways. Using the rule in (11) with this threshold, all homography matrices that are within the expected range according to the aforementioned physical restrictions are accepted for updating the estimation, and the rest are rejected. A similar procedure must be followed to fix the threshold t_n for any other application environment, taking into account its particular kinetic restrictions.

As a result of the Kalman correction stage, an stable and reliable estimate, $\bar{\boldsymbol{H}}^{\mathrm{e}}$, is obtained for the normalised homography. The normalisation is undone by a transformation complementary to (4), that is, $\boldsymbol{H}^{\mathrm{e}} = \boldsymbol{K}\bar{\boldsymbol{H}}^{\mathrm{e}}\boldsymbol{K}^{-1}$. Alignment of the current and previous images is achieved by warping the latter with $\boldsymbol{H}^{\mathrm{e}}$.

An example of image warping with a planar homography is shown in Fig. 4. In the first row the previous and the current frame are displayed; below, the left column shows the warping of the previous image with the instantaneous homography $\boldsymbol{H}^{i}$, whereas the right column corresponds to the warping with $\boldsymbol{H}^{\mathrm{e}}$. As can be observed, if the image is warped by the instantaneous homography $\boldsymbol{H}^{i}$, as done in many approaches in the literature, the resulting image features a slight deformation compared to the real image (see Fig. 4*c*). Conversely, with the proposed method a robust estimation, $\boldsymbol{H}^{\mathrm{e}}$, of the homography is obtained, and hence the aligned image (see Fig. 4*d*) is very similar to the current image in the ground region.

3.3 Ground plane region detection

As stated in the foundation of the method, the points belonging to the ground plane are expected to have coherent motion patterns. Namely, all the static elements in the ground plane are projected from the previous to the current frame through $\boldsymbol{H}$ as in (1). Conversely, all the background elements (that are not on the ground plane) and the moving objects are subject to different transformations. Hence, the difference between the current image and the previous image warped with the estimated homography $\boldsymbol{H}^{\mathrm{e}}$ is expected to be null for the static elements on the ground plane and non-zero for the rest of the image. Therefore the differences between aligned images delimitate the ground plane region. These differences are detected evaluating sum of absolute differences (SAD) between the current image and the previous image warped. Fig. 4*e* shows an example of the difference between the current frame (Fig. 4*b*) and the previous image warped (Fig. 4*d*). Note that moving objects (i.e. vehicles) are clearly distinguished by white horizontal patches that appear in their contact zones with the ground plane.

Fig. 5 shows some examples of road region detection for traffic scenarios. To obtain road regions, images are scanned bottom to up in search of pixels with significant SAD value, which correspond to moving objects or to elements above the ground plane. The regions above these pixels do not belong to the road plane. Note as well that the projection of the road in the image is bounded by the vanishing line (i.e. the horizon) [21]. Hence, the images in Fig. 5 are constructed by overlaying the input image below the regions of significant difference within the region of interest (ROI; below the vanishing line) with a grey mask. As shown, background elements and moving objects are located out of the road plane region. Note that ground plane estimation is consequently a powerful basis for the attainment of object detection and tracking. Namely, the elements out of the ground plane may be further examined; in particular, different images of a sequence may be analysed in order to locate these elements that do not belong to the ground plane and that additionally have a shape or motion pattern expected from a moving object.

4 Experiments and discussion

This section aims at showing the robustness of the proposed method, and its applicability to object detection. Hence, an approach is addressed here for vehicle detection and tracking in order to better assess the performance of the method on the traffic environment, both visually and statistically. The straightforward approach for object tracking is based on Kalman filtering. In effect, the motion of vehicles is expected to be smooth over the ground plane, hence some correlation is expected between vehicle

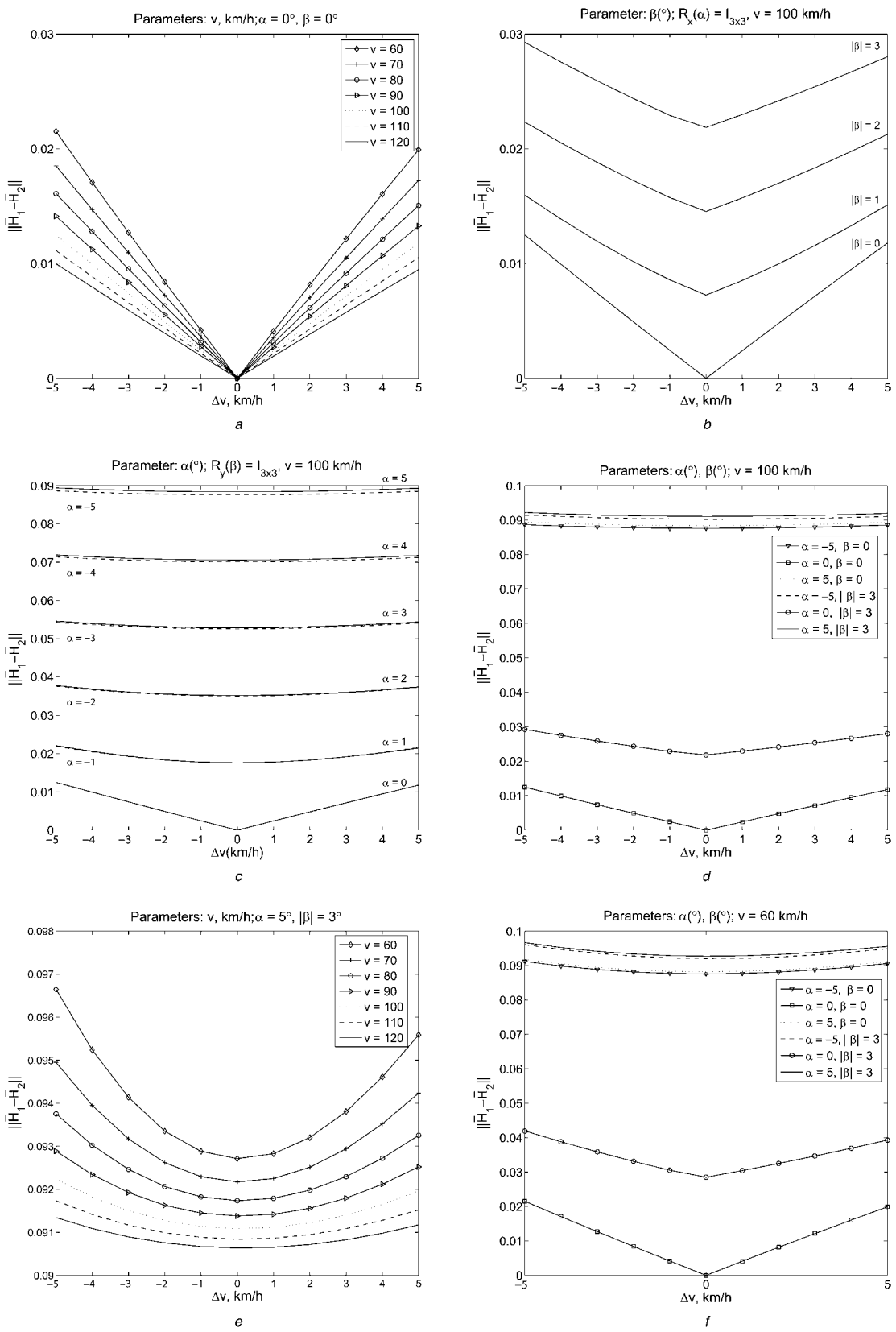

Figure 3 *Analysis of* $\|\bar{\boldsymbol{H}}_1 - \bar{\boldsymbol{H}}_2\|$ *as a function of the speed of the vehicle, v, its variation,* Δv*, and the rotation angles,* α *and* β

a $\alpha = \beta = 0^\circ$, *v* is a parameter
b $v = 100$ km/h, $\alpha = 0^\circ$; β is a parameter
c $v = 100$ km/h, $\beta = 0^\circ$; α is a parameter
d $v = 100$ km/h, α and β are parameters
e $\alpha = 5^\circ$, $|\beta| = 3^\circ$; *v* is a parameter
f $v = 60$ km/h; α and β are parameters

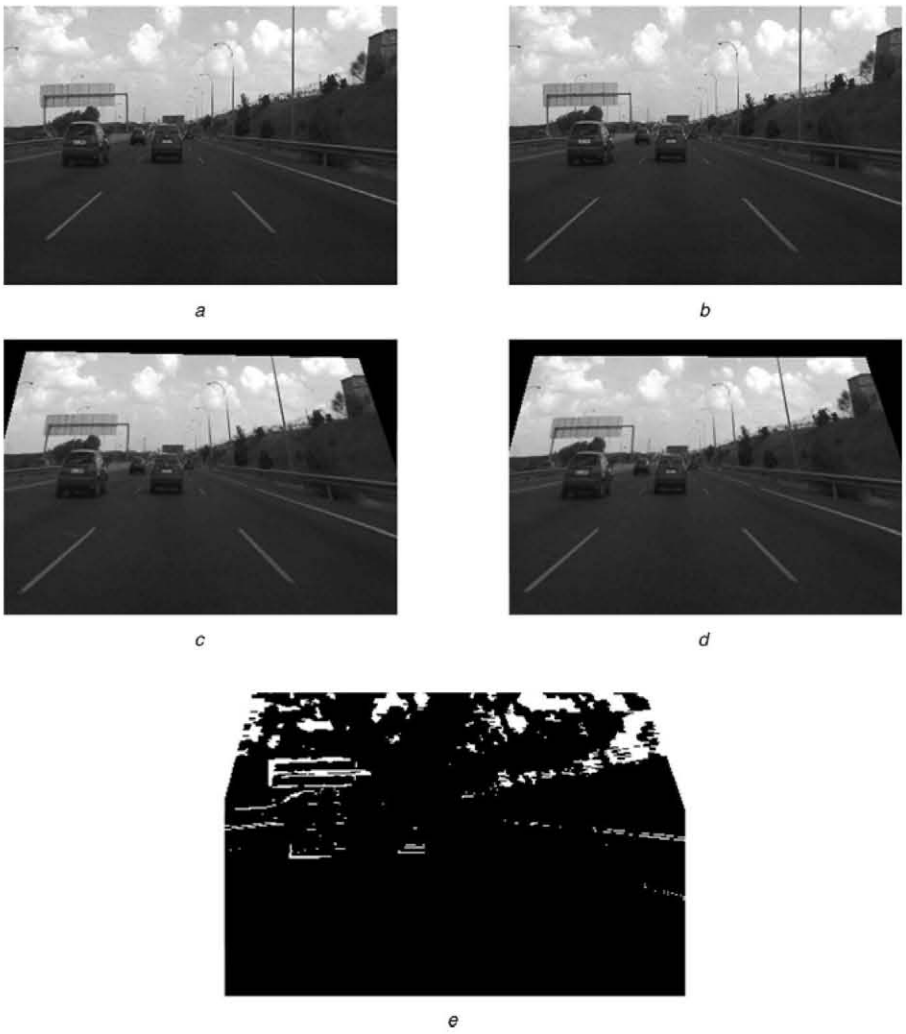

Figure 4 *Example of homography applied to align images*

a Previous frame
b Current frame
c Previous frame warped with inaccurate instantaneous homography $\boldsymbol{H}^{\mathrm{i}}$
d Previous frame warped with robustly estimated homography $\boldsymbol{H}^{\mathrm{e}}$
e Difference between images *b* and *d*

positions in successive images. Therefore a linear model can be assumed for the evolution of the vehicle position, with a locally constant velocity. In this context, the bird's eye view of the road plane, given by the inverse perspective mapping (IPM) or plane rectification through image warping [22], can be used to model the linear motion of the vehicles, as explained in [23]. In this domain, the dynamic and observation models of the system can be written, respectively, as

$$\begin{aligned} \boldsymbol{s}_k &= \boldsymbol{F}\boldsymbol{s}_{k-1} + \boldsymbol{n}_{k-1} \\ \boldsymbol{m}_k &= \boldsymbol{T}\boldsymbol{s}_k + \boldsymbol{u}_k \end{aligned} \tag{13}$$

$$\boldsymbol{F} = \begin{pmatrix} I_2 & \Delta t \cdot I_2 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_2 \end{pmatrix}, \quad \boldsymbol{H} = \begin{pmatrix} I_2 & 0 & 0 \\ 0 & 0 & I_2 \end{pmatrix}$$

where the state vector $\boldsymbol{s}_k$ is composed of the information regarding each vehicle, that is, its position (x_1, x_2), velocity $(\dot{x}_1, \dot{x}_2)$, width (w), and height (h) as in [23]

$$\boldsymbol{s}_k = (x_1, x_2, \dot{x}_1, \dot{x}_2, w, h)^{\top} \tag{14}$$

the measurement vector $\boldsymbol{m}_k$ is composed of the observations at time k of the vehicle position and dimension, and the noise distributions $\boldsymbol{n}_k$ and $\boldsymbol{u}_k$ are independent and Gaussian. The position and width of the vehicles are obtained from the analysis of the road region estimation at each instant. In particular, they are given by the regions that do not belong to the road, that is, regions which have a significant SAD value, as imposed in Section 3.3. These regions might as well correspond to elements in the background, hence only the regions that have a size, shape and motion coherent to that expected from vehicles are taken into consideration. Additionally, using the update stage of the Kalman filter, the prediction of each object is

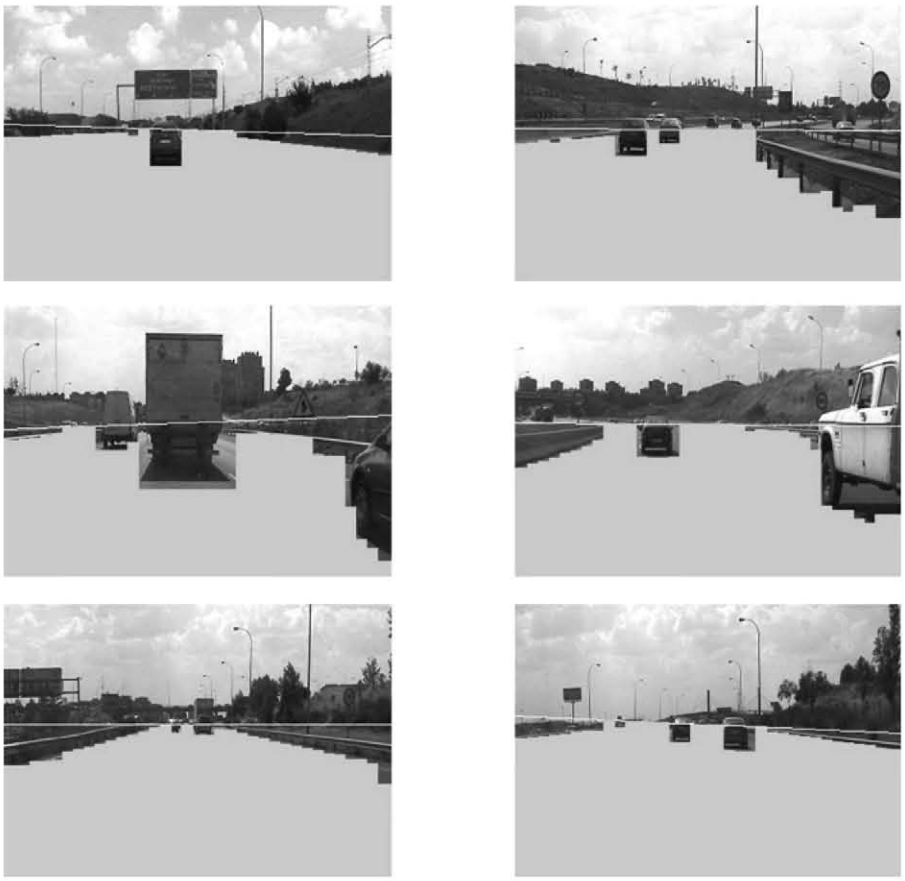

Figure 5 *Road plane detection: regions of the input image that correspond to the road region are overlaid with a grey mask*

corrected with the new measurement obtained for it. Using this approach, vehicle detection and tracking is achieved, hence there is correlation between vehicles at different points in time. Therefore it is possible to know the trajectory (e.g. direction) of the vehicles, which is of high interest, for instance to analyse drivers behaviour.

The proposed method for ground plane estimation followed by an object tracking strategy as described above has been tested for the traffic scenario with different driving conditions. The test sequences were acquired using a SONY HDR-HCR5E camera with a resolution of 360×288 mounted on an on-board platform. A general purpose PC working at 2 GHz is employed for image processing, rendering an output frame rate of around 10 fps. In order to statistically assess the performance of the method, the vehicle true positive, false positive and false negative detection rates have been evaluated for different test sequences with a total duration of 33 min. The true positive rate is defined as the number of correctly detected vehicles over the total number of detectable vehicles. The false negative rate is immediately derived from the previous parameter as the rate of non-detected vehicles. The false positive rate is defined as the number of false positives (i.e. regions that are classified as vehicles by the system and which are actually not vehicles) over the total number of detectable vehicles. A vehicle is considered detectable since it enters the ROI until it abandons the ROI. In turn, the ROI comprises the own and the two adjacent lanes, and is limited to a maximum longitudinal distance d that depends on the camera calibration. In addition, if there is a vehicle (partially or totally) occluding others, only the occluding vehicle is considered as detectable. A vehicle is considered to be correctly detected when it is tracked by the system in at least 90% of its existence time within the ROI (a small number of losses is admitted owing to the intrinsic limitations of the bird's-eye view given by the IPM, that is, blind regions in the near area, and inaccuracies in the upper part due to non-perfect plane rectification).

Test sequences involve a set of realistic scenarios, including different weather (cloudy/sunny) and traffic load (low/heavy traffic) conditions. Table 1 summarises the rates obtained for the different scenarios, as well as the absolute figures for each scenario, that is, number of true positives, number of false positives, and total number of vehicles. Using the relatively simple object detection strategy described above, we obtain a mean vehicle detection rate of 90.3% and a false positive rate of 8%, that demonstrates the potential of the proposed method. As expected, the system provides the best detection results, 97.5%, for cloudy sequences with low traffic density (in effect, under these conditions road

Table 1 Detection results for different scenarios

	Type of scenario				Total
	Cloudy/low traffic	Cloudy/heavy traffic	Sunny/low traffic	Sunny/heavy traffic	
time	6′32″	6′54″	10′14″	10′04″	33′44″
no. of true positives	40	79	70	82	271
no. of false positives	3	6	6	9	24
no. total vehicles	41	84	77	98	300
true positive rate (%)	97.5	94.0	90.9	83.7	90.3
false positive rate (%)	7.3	7.1	7.8	9.2	8.0
false negative rate (%)	2.5	6.0	9.1	16.3	9.7

features are less likely to be cluttered, and illumination is more homogeneous). False positive rates are near 8% regardless of the scenario, as they are mostly produced due to elements on the road side, such as guard rails.

Some example results are shown in Fig. 6. The slight position error in the lower bound is due to the differential nature of the method (i.e. the difference starts at the predicted position of the vehicles rather than at their actual position). This slight error could be compensated by calculating the velocity of the vehicles. This can be inferred by first taking from the controller area network the velocity of the vehicle in which the camera is mounted, and then adding to it the relative velocity of the different target vehicles obtained from their tracking process. With the estimated absolute velocity of the vehicles, if camera calibration is available and frame rate is known, the space covered by the vehicles can be estimated and, via the camera calibration, the corresponding pixels in the image can be compensated.

The strength of the method presented lies on the robustness of the defined homography prediction model. The rule to

Figure 6 *Examples of moving vehicle detection after road region detection*

classify the new observations and thus update the homography prediction depends on the norm of the difference between the new homography measurement and the predicted one. This rule ensures that those measurements that are compliant with the kinetic restrictions of the vehicles are accepted. Fig. 7 shows that the difference of the norm is indeed significant to detect erroneous measurements. In Fig. 7*a* the evolution of the norm of the instantaneous homography matrix in time is plotted for a test sequence. As can be observed, the norm of the homographies rejected by the designed rule (which are marked with black circles) is in general very different to that of the accepted ones. A detail of Fig. 7*a* around the accepted measurements is given in Fig. 7*b*. As can be observed only two of the rejected homographies have a norm similar to that of the correct measurements. The norm of the predicted homography for every time point has also been displayed in the figure. Note that the prediction of the homography adapts smoothly to the new measurements.

The difference between the instantaneous and the predicted homographies is plotted in Fig. 7*c* and, zoomed near the origin, in Fig. 7*d*. The dashed line in Fig. 7*d* depicts the threshold for homography acceptance. All the strong outliers in Fig. 7*a* are filtered in effect by the proposed rule (the black squares in Fig. 7*c* corresponding to rejected measurements are in the same time points as the black dots in Fig. 7*a*). These must have been generated due to wrong correspondences in the equation system derived from (2), since they are out of the range given by the kinetic restrictions associated to the moving platform. Remarkably, in the example, the norm of the matrix difference allows to identify two outliers that have a norm similar to that of the predicted homography, which are actually produced by wrong correspondences, and therefore rejected.

Besides, one of the main advantages of the proposed method is that due to its predictive nature, the method is able to perform for long periods of time without new measurements. In fact, a prediction of the homography is available at every time point; hence, this prediction (instead of the erroneous or non-existing instantaneous measurement) can be used to achieve image alignment and eventually object detection. This can be observed in Figs. 7*a* and *c*, in which no new measurement of the homography exists for long stretches (e.g. frames 152–253, 578–734). During these periods, the prediction is relied on until new measurements are provided, with no impact on the operation of the system.

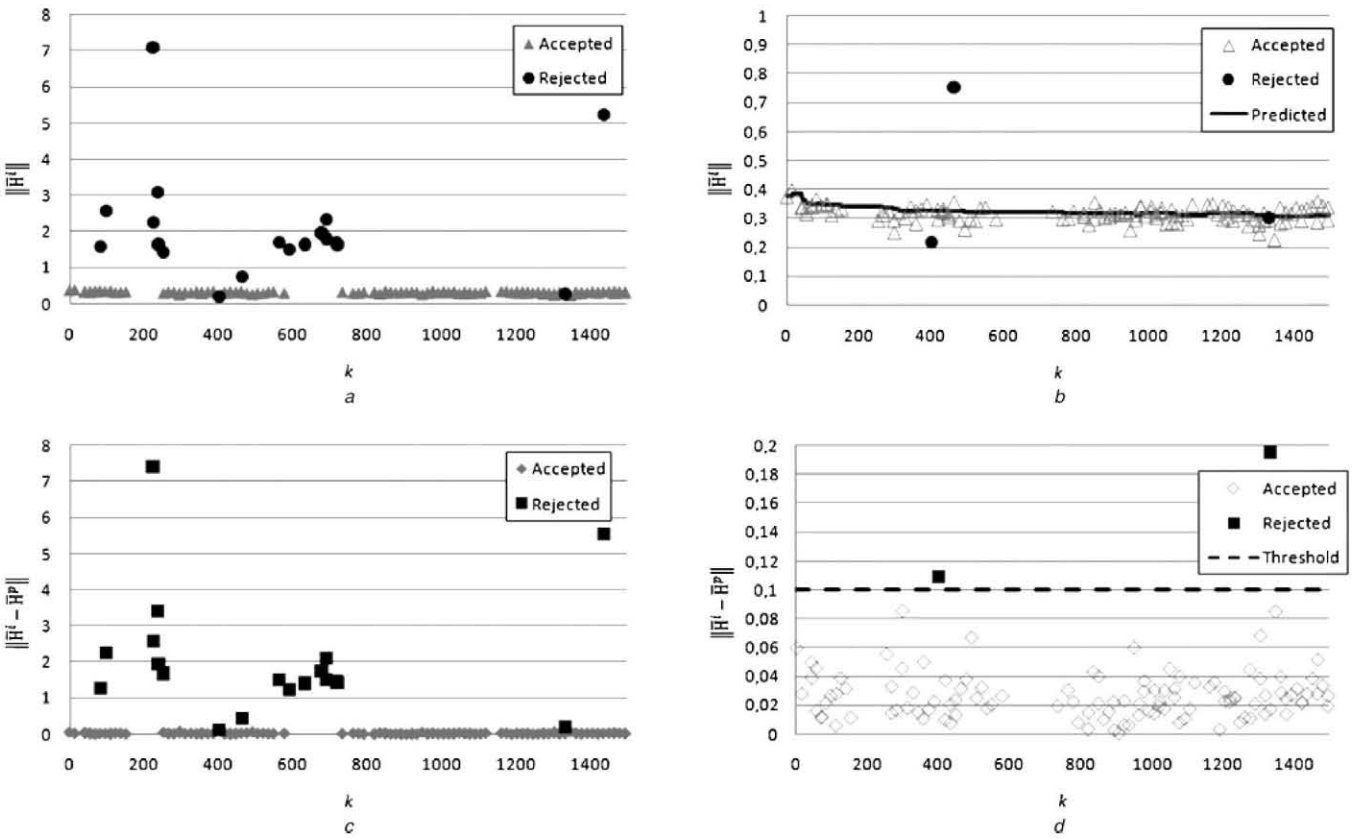

Figure 7 *Evolution of* $\|\bar{H}^i\|$ *and* $\|\bar{H}^i - \bar{H}^p\|$ *for a test sequence*

a Evolution of $\|\bar{H}^i\|$
b Figure *a* zoomed near the origin
c Evolution of $\|\bar{H}^i - \bar{H}^p\|$
d Figure *c* zoomed near the origin

5 Conclusions and future work

In this paper, a method for estimating the ground plane in a dynamic scene captured by a non-stationary camera has been presented. The proposed approach computes the homography of the ground plane for two consecutive images to achieve image alignment. As opposed to other typical approaches, this method does not entail explicit computation of camera motion nor 3D ground plane parameter estimation. Alternatively, it is based solely on feature matching across successive images and a new homography calculation framework.

The planar homography is assumed to be locally stable and change smoothly. Based on this assumption, the homography calculation has been modelled as a linear data estimation problem. The method is especially suited for traffic environments, in which the scarcity of feature points and the instability of the camera are bound to render imprecise or erroneous instantaneous measurements. With the proposed time-filtering method, although instantaneous homography measurements may be inaccurate, a robust estimate of the homography is attained at every time point.

A reliable ground plane estimation is essential for the detection of objects moving on the ground plane. The potential of the method has been shown by complementing the method with an object detection strategy. This yields remarkable results even if a rather straightforward strategy is used, which proves the robustness and applicability of the proposed method for road region estimation. Future work will focus on the study of more sophisticated object tracking strategies, such as particles filters or extended Kalman filters (EKF). Owing to its non-linear nature, particle filter allows one to perform the analysis on the original image, where vehicles dynamics are non-linear because of perspective effect, thus avoiding the need for an IPM. In addition, it enables more complex observation models than Kalman filter. Alternatively, the use of a second-order EKF will also be explored for vehicle tracking in the transformed domain as an enhancement of the proposed Kalman filtering. Naturally, the use of sophisticated object tracking strategies, as those suggested above, upon the proposed method for ground plane region estimation is expected to provide more precise and complete results.

6 Acknowledgments

This work has been supported by the Ministerio de Educación y Ciencia of the Spanish Government under projects TEC2007-67764 (SmartVision) and TEC2006-26845-E (HIGHWAY); by the Ministerio de Ciencia e Innovación and co-financed by the Fondo Europeo de Desarrollo Regional FEDER under project PSE-370000-2009-009 (TECMUSA).

7 References

[1] YAMAGUCHI K., WATANABE A., NAITO T.: 'Road region estimation using a sequence of monocular images'. Proc. Int. Conf. on Pattern Recognition, Tampa, USA, December 2008, pp. 1–4

[2] ZHOU J., LI B.: 'Homography-based ground detection for a mobile robot platform using a single camera'. Proc. IEEE Int. Conf. on Robotics and Automation, Orlando, USA, May 2006, pp. 4100–4105

[3] ZHOU H., WALLACE A.M., GREEN P.R.: 'A multistage filtering technique to detect hazards on the ground plane', *Pattern Recognit. Lett.*, 2003, **24**, (9), pp. 1453–1461

[4] CHUMERIN N., VAN HULLE M.M.: 'Ground plane estimation based on dense stereo disparity'. Proc. Int. Conf. on Neural Networks and Artificial Intelligence, Minsk, Belarus, May 2008, pp. 209–213

[5] SIMOND N.: 'Reconstruction of the road plane with an embedded stereorig in urban environments'. Proc. IEEE Intelligent Vehicles Symp., Tokyo, Japan, June 2006, pp. 70–75

[6] STEIN G.P., MANO O., SHASHUA A.: 'A robust method for computing vehicle ego-motion'. Proc. IEEE Intelligent Vehicles Symp., Dearborn, USA, October 2000, pp. 362–368

[7] BLANCO J.-L., GONZALEZ J., FERNANDEZ-MADRIGAL J.-A.: 'Mobile robot ego-motion estimation by proprioceptive sensor fusion'. Proc. Int. Symp. on Signal Processing and its Applications, Sharjah, UAE, February 2007, pp. 1–4

[8] CAO Y., COOK P., RENFREW A.: 'Vehicle ego-motion estimation by using pulse-coupled neural network'. Proc. Int. Machine Vision and Image Processing Conf., Maynooth, Ireland, September 2007, pp. 185–191

[9] GAVRILA D.M., MUNDER S.: 'Multi-cue pedestrian detection and tracking from a moving vehicle', *Int. J. Comput. Vis.*, 2007, **73**, (1), pp. 41–59

[10] ESS A., LEIBE B., SCHINDLER K., VAN GOOL L.: 'Robust multiperson tracking from a mobile platform', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (10), pp. 1831–1846

[11] RIDDER C., MUNKELT O., KIRCHNER H.: 'Adaptive background estimation and foreground detection using Kalman-filtering'. Proc. Int. Conf. on Recent Advances in Mechatronics, Istanbul, Turkey, August 1995, pp. 193–195

[12] KOLLER D., WEBER J., MALIK J.: 'Robust multiple car tracking with occlusion reasoning'. European Conf. on Computer Vision, 1994 (*LNCS*, **800**), pp. 189–196

[13] KOLLER D., DANILIDIS K., NAGEL H.-H.: 'Model-based object tracking in monocular image sequences of road traffic scenes', *Int. J. Comput. Vis.*, 1993, **10**, (3), pp. 257–281

[14] HARTLEY R.I., ZISSERMAN A.: 'Multiple view geometry in computer vision' (Cambridge University Press, 2000, 1st edn.)

[15] NIETO M., SALGADO L., JAUREGUIZAR F., CABRERA J.: 'Stabilization of inverse perspective mapping images based on robust vanishing point estimation'. Proc. IEEE Intelligent Vehicles Symp., Istanbul, Turkey, June 2007, pp. 315–320

[16] LOWE D.G.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110

[17] WELCH G., BISHOP G.: 'An introduction to the Kalman filter'. Tech. Report TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 2004

[18] GIBBONS J.D., CHAKRABORTI S.: 'Nonparametric statistical inference' (Marcel Dekker, 1971, 4th edn.)

[19] MOON T.K., STIRLING W.C.: 'Mathematical methods and algorithms for signal processing' (Prentice Hall, 1999, 1st edn.)

[20] American Association of Highway and Transportation Officials: 'A policy on geometric design of highways and streets' (Washington, DC, 1984, 5th edn. 2004)

[21] CRIMINISI A., REID I., ZISSERMAN A.: 'Single view metrology', *Int. J. Comput. Vis.*, 2000, **40**, (2), pp. 123–148

[22] BERTOZZI M., BROGGI A.: 'Gold: a parallel real-time stereo vision system for generic obstacle and lane detection', *IEEE Trans. Image Process.*, 1998, **7**, (1), pp. 62–81

[23] ARRÓSPIDE J., SALGADO L., NIETO M., JAUREGUIZAR F.: 'Real-time vehicle detection and tracking based on perspective and non-perspective space cooperation'. Proc. SPIE Int. Conf. on Real-Time Image and Video Processing, San Jose, 7244H, January 2009, pp. 1–12