the previous and current views of the scene. In particular, the homography (i.e. projective transformation) of the road plane is estimated between these two points in time. This allows us to generate a prediction of the road plane appearance in future instants. However, vehicles (which are generally the only objects moving on the road plane) feature inherent motion in time, hence their projected position in the plane differs from that observed. The regions involving motion are identified through image alignment of the current image and the previous image warped with the homography. These regions will correspond to vehicles with high probability.

5.2.1 Homography calculation The first step towards image alignment is the calculation of the road plane homography between consecutive frames. As shown in [37] the homography that relates points of a plane between two different views can be obtained from a minimum of four feature correspondences by means of the Direct Linear Transformation (DLT). Indeed, in many applications the texture of the planar object allows to obtain numerous feature correspondences using standard feature extraction and matching techniques, and to subsequently find a good approximation to the underlying homography. However, this is not the case in traffic environments: the road plane is highly homogeneous, and hence most of the points delivered by feature detectors applied on the images belong to background elements or vehicles, and few correspond to the road plane. Therefore, the resulting dominant homography (even if using robust estimation techniques) is in general not that of the road plane.

To overcome this problem, we propose to exploit the specific nature of the environment. In particular, highways are expected to have different kind of markings (mostly lane markings) painted on the road. Therefore, we propose to first use a standard lane marking detector (such as the ones described in [33–35]) and then to restrict the feature search area in extended regions around lane markings. Nevertheless, the resulting set of correspondences will still typically be scarce, and some of them may be incorrect or inaccurate, depending on the sharpness of the lane marking corners and on the resolution of the image around them. Hence, the instantaneous homography computed from feature correspondences using DLT might be highly unreliable (errors in one of the points will have a large impact in the solution to the equation system of DLT), and sometimes the number of points is not even sufficient to compute it.

For the above-mentioned reasons, intermediate processing of the instantaneous homography is necessary. This is achieved in the present work by means of a linear estimation process based on Kalman filtering. Let us first inspect the analytical expression of the homography between two consecutive instants. Fig. 3 illustrates the situation of a vehicle with an on-board camera moving on a flat road plane, \( \pi_0 = (n^\top, d)^\top \), where \( n = (0, 1, 0)^\top \) and \( d \) is the distance between the camera and the ground plane. The coordinate system of the camera at time \( k_1 \) is adopted as the world coordinate system, where \( z \) axis indicates the driving direction. At time \( k_2 \) the camera has moved to position \( C_2 \), and rotation \( R_y(\alpha) \) might have occurred around the \( x \)-axis due to camera shaking (\( \alpha \) denotes the change in the pitch angle). Additional rotation \( R_y(\beta) \) models variations in the yaw angle (i.e., around the \( y \)-axis), which must be considered in the case the vehicle changes lanes or takes a curve. From the previous discussion, and assuming a pinhole camera model, the camera projection matrices at times \( k_1 \) and \( k_2 \) are respectively

\[
P_{1} = K[I|0] \\
P_{2} = KR_x(\alpha)R_y(\beta)[I - C_2]
\]

(20)

The homography \( H \) relates the projections, \( x_1 \) and \( x_2 \), of a 3D point, \( X \in \pi_0 \), in two different images. Its expression can be derived from the equations in (20). In effect, for the first view it is \( x_1 = P_1X = K[I|0] \) and hence any point in the ray \( X = (x^2_1(K^{-1})^\top, \rho)^\top \) projects to \( x_1 \). The intersection of this ray and the plane \( \pi_0 \) determines the value of the parameter \( \rho \); it is \( \pi_0^\top X = n^\top K^{-1}x + d\rho = 0 \), and thus \( \rho = -n^\top K^{-1}x_1/d \). The projection \( x_2 \) of the point \( X \) into the second view is

\[
x_2 = P_2X = KR_x(\alpha)R_y(\beta)[I - C_2]X = \\
= KR_x(\alpha)[R_y(\beta) - R_y(\beta)C_2](x^2_1(K^{-1})^\top, \rho)^\top = \\
= KR_x(\alpha)[R_y(\beta)K^{-1}x_1 + t\rho] = \\
= KR_x(\alpha)[R_y(\beta) - tn^\top/dK^{-1}x_1]
\]

where \( t = -R_y(\beta)C_2 \). This vector constitutes the translation in the direction in which the vehicle is heading and is thus given by \( t = (0, 0, 1)^\top v/f_r \), where \( v \) is the velocity of the vehicle and \( f_r \) is the frame rate. From the above equations the expression of the homography of the plane \( \pi_0 \) between \( k_1 \) and \( k_2 \) is derived:

\[
H = KR_x(\alpha)[R_y(\beta) - tn^\top/dK^{-1}]
\]

(21)