What can be done with an embedded stereo-rig in urban environments ?

Nicolas SIMOND [a] , Patrick RIVES [b]

[a] *Centre de Robotique, Ecole des Mines de Paris, 60 Bd Saint-Michel, 75 272 Paris Cedex 06, France*

[b] *Projet ICARE, INRIA Sophia-Antipolis, 2004 route des Lucioles, 06 902 Sophia-Antipolis Cedex, France*

Abstract

Key words: stereo-vision, mobile robotics, urban environment, super-homography, trajectography, 2.5D reconstruction
PACS: I.2.9.0: Autonomous vehicles I.2.10.0: 3D/stereo scene analysis I.2.10.4: Motion I.4.6.0: Edge and feature detection

1 Introduction

A prerequisite to the design of future Advanced Driver Assistance Systems for cars is a sensing system providing all the information required for high-level driving assistance tasks. In particular, such a sensing system should be able to provide a robust and accurate real time localization in constrained urban environments. Moreover, if we want to address Autonomous Guided Vehicles (AGVs) issues, we must also take into account on board sensors able to provide representations of the environment supporting obstacle-free trajectory planning algorithms and controller design based on reactive sensor-based control tasks.

Commonly, the localization in outdoor environment is based on the fusion between GPS data and data provided by proprioceptive sensors embedded on the vehicle like odometers and/or Inertial Navigation System (INS). In urban

[1] This work was funded by the European project Cybercars. The authors thanks the IMARA (INRIA Rocquencourt) and LIVIC (INRETS-LCPC) projects which provide us the stereo-sequences we use.

environment, this process is enhanced using a prior knowledge from maps thanks to a map matching algorithm which constraints the trajectory of the vehicle to belong to the road. Nevertheless, the success of such techniques is closely related to the reliability of the GPS data and its availability during the navigation task. Practically, in a dense urban environment (ie. old city downtown), these prerequisites hold rarely. Due to the presence of vertical structures in the architecture, the visibility of a sufficient number of satellites for positioning (at least 4) is time dependent and the signals are frequently corrupted by multipaths propagation. When reliable GPS data are missing, the localization process is only supported by a dead reckoning method based on the integration of odometry data provided by internal proprioceptive sensors. It is well-known that such sensors are subject to an important drift leading to a biased localization.

On the other hand, it is now clearly admitted that, in a near future, every car will be equipped with on-board exteroceptive sensors like vision, lidar or radar for ADAS applications. Such sensors are well-adapted to capture the interactions between the vehicle and the *local* environment in the vehicule-centered frame. Under certain assumptions, from these data, it is possible to simultaneously estimate the local structure of the environment and the ego-motion of the vehicle. This problem often called *Simultaneous Localization and Mapping (SLAM)* problem was intensively studied in the case of indoors robotics. Nevertheless, the transposition of SLAM indoors methods to applications in urban environments is not straightforward due to the complexity and the dynamical character of the urban context.

SLAM algorithms for specific inner-city applications have been developed for the last five years. To ego-locate from the cluttered environment, the mobile robots have first to segment out the moving objects before building maps and localizating task. Hence, the CMU Navlab purposed in 2003 to accomplish *SLAM with Detection And Tracking of Moving Objects (DATMO)* [33] that became in a unified approach *SLAM and Moving Object Tracking (SLAM-MOT)* [32], in 2004. When a model of the environment is known, the vehicle localization tends to a *Structure From Motion (SFM)* problem ([22]) because opposite to SLAM methods, the duration of the environment visibility is very short with a travelling vehicle. Recently, collaborative GPS/INS navigation ([4]) is investigated to provide data localization to a vehicle in a GPS shading area.

Basically, on-board vision was implemented to improve the robustness of the navigation task by analyzing the evolution of the environment. Vision has in fact many applications on obstacle detection and lateral control of a path following limited by boundaries (line or lane tracking). Nevertheless, many authors try to estimate the ego-motion of the camera(s) assuming a prior knowledge on the environment:

- optic flow ([31,29]) is used in case of unstructured environment or textureless environment,
- direct method ([28,15]) provides reliable results assuming a calibrated camera and a motion model reduced to the three main parameters,
- homography ([19]) to detect planes in the environment and compute then the relative motion of the camera.

Note that the three last references allow the estimation of a vehicle ego-motion following a textureless road.

Due to the fact that urban environment are structured, the localization of a vehicle is affordable assuming a transposition of visual methods dedicated to indoor environment. The relative camera orientation can be computed from the location of vanishing points ([1,11]). On the opposite way, the navigation of a mobile robot requires the fusion of data from odometer, INS, GPS, vision and an a priori knowledge of the static environment as a Geographical Information Systems (GIS) to improve the reliability of the global localization with identifying "natural" beacons like [10] of the NYU Avenue Project. The introduction of artificial beacons in a city is thus not an acceptable solution due to the scale problem. Authors prefer identifying vertical edges which are precisely referenced in GIS ([6,13]). Another way consists on the update of the current view with a set of images, recorded and geo-referenced during a calibration phase ([14,23]).

We present in this article a novel method to estimate the trajectory of a vehicle navigating in an urban environment thanks to the data provided by an onboard stereo-vision system. We assume a canyon-like environment which contains some static planes (road, frontages) where characteristic features (points, lines) can be extracted. The purposed method copes with the dense traffic conditions: the free space required (first ten meters in front of the vehicle) is slightly equivalent to the security distance between two vehicles. Moreover, the estimation of the pose of the stereo rig along the navigation path allows us to build a polyhedral model of the environment using a SLAM technique.

The paper is organized as follows. In the first section, a simplified model of the urban environment is proposed and we present an overview of the method. The second section focuses on the segmentation of the main planes along the sequence of images. The third section is devoted to the robust estimation of the coplanar features which belong to the road plane from the sequences of images provided by the stereo rig. In the fourth section, we develop a method based on the decomposition of the homographies linking the different views in order to estimate the ego-motion and the geometry of the planes in the scene. Finally, thanks to a mosaicing-like technique, we compute both the trajectory of the vehicle and a representation of the underlying road. Experimental issues from real data are presented and discussed. We conclude this article in the last

section by proposing perspectives.

2 Problem Statement

Viewed from inside the vehicle, we can consider an urban street as a kind of canyon usually bounded by three main surfaces: the road and two others formed by the frontages in each side of the road. Each of them will be assumed to be locally planar in the few ten meters inside the field of view observed by the cameras in front of the vehicle. Furthermore, the frontages are formed with sets of quasi vertical planes containing patterns whose edges are generally aligned with the two main directions corresponding to the road axis and the vertical axis. Concerning the road signs lying on the ground, they are generally aligned with the longitudinal and the transversal road axis. We consider the vehicle is equipped with an on-board stereo-rig providing standard grey level images at video rate. The vehicle does not exceed the speed limit in urban environment (abroad 50 km/h in European countries). Consequently, we also assume small displacements of the features between two successive images in the sequence.

2.1 Mathematical prerequisites

We denote the discrete time index by the variable k. At each iteration (k), two $[n_{raw} \times n_{col}]$ images $\mathcal{I}_l^k$ and $\mathcal{I}_r^k$ (respectively, left and right), are recorded. Let $\mathbf{P_s}$ be a 3D point of the environment, its projection onto the image is represented by a $[3 \times 1]$ homogeneous vector $\mathbf{p_s^k} = [u, v, 1]^t$. The contour $\mathbf{L_s}$ in the 3D environment projects onto the image as a $[3 \times 1]$ homogeneous vector $\mathbf{l_s^k} = [\cos(\theta), \sin(\theta), -\rho]^t$ where θ and ρ are the polar representation of the line in the image frame.

A 3D point in the scene is related to its projection in the image by a projective equality: $\mathbf{p_s} \propto \mathsf{K}[\mathsf{R},\mathbf{t}].\mathbf{P_s}$ where K represents the $[3 \times 3]$ matrix of intrinsic parameters, R and $\mathbf{t}$ the rotation matrix and the translation vector between the scene and the camera frames. The variable q will be introduced to represent the evolution of an iterative process.

Let us consider a set of features lying on a same plane in the scene and the images of these features taken by several cameras from different points of view (Fig.1). It is now wellknown [7,12] that for each couple of images $\mathcal{I}_a$ and $\mathcal{I}_b$ it exists a projective homography which links the coordinates $\mathbf{p_a}$ and $\mathbf{p_b}$ of the features in the images $\mathcal{I}_a$ and $\mathcal{I}_b$ such that $\mathbf{p_b} \propto \mathsf{H_{ba}}\mathbf{p_a}$. In the Fig.1, homographies induced by the π plane between the stereo images are noted

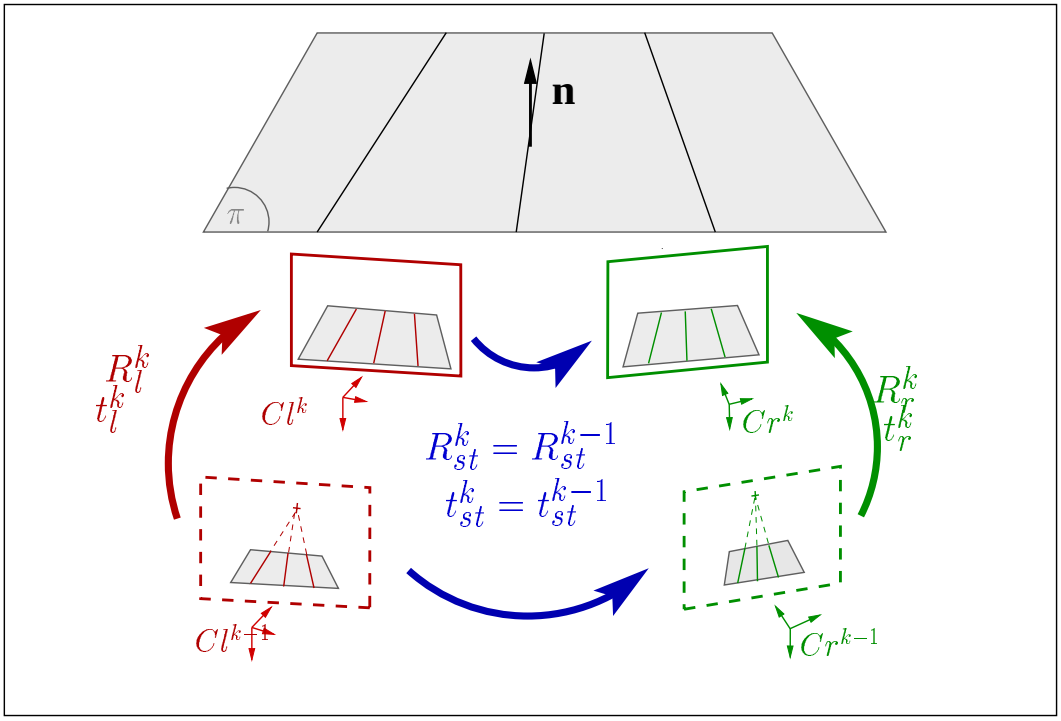


Fig. 1. *Rotation* (R) *and translation* **t** *motions between the discrete time (k-1) and (k), induced by the stereo-rig motion above the plane* π *which is defined with the normale to plane* **n**.

$\mathsf{H_{st}}$, whereas homographies between consecutive left images (resp. right) are noted $\mathsf{H_l}$ (resp. $\mathsf{H_r}$).

2.2 Overview of the Method

Among the different planes constituting the urban canyon, the road plane takes a central importance. Sometimes, in large avenues or in presence of trees, it will be the unique plane which remains visible when the vehicle is moving. So, it is fundamental for navigation and guidance purposes to robustly and accurately detect and characterize the road plane. Unfortunately, due to the uniformity of the surface, the lack of texture and the few of characteristic features (excepted the road signs), this detection will be the most difficult. In the next two sections, we detail how we can estimate the ego-motion of the vehicle using the features extracted from the road plane only. The results will be extended to the others planes when they are available, in the fourth section.

The basic idea of the method is to take advantage both the spatial constraints linking the left and right cameras to extract the features and the temporal constraints between the successive views. An overview of the complete algorithm is provided in the Fig. 2.

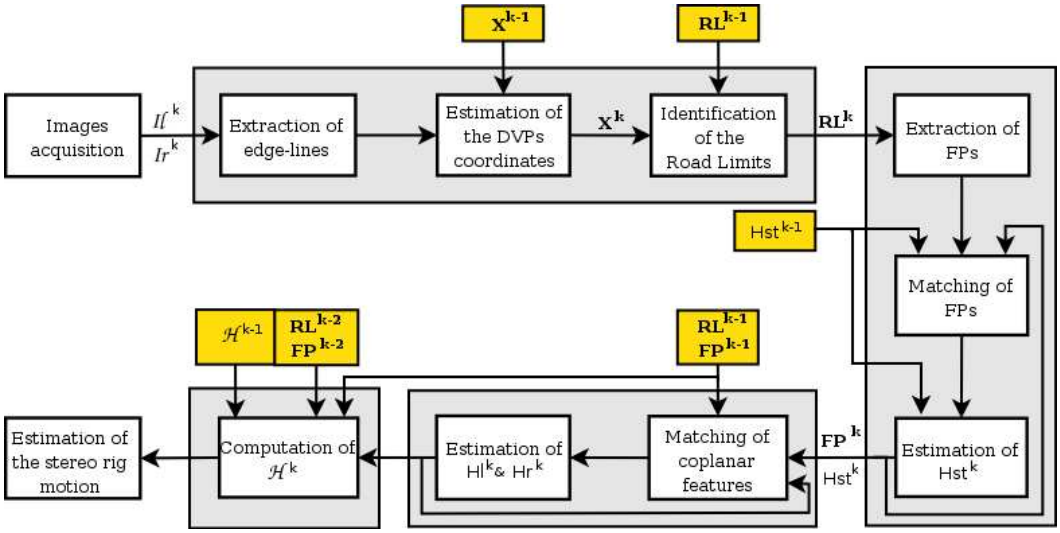


Fig. 2. *Overview of the method. The first stage corresponds to the segmentation of the road plane into the images* $\mathcal{I}_l^k$ *and* $\mathcal{I}_r^k$*. In the second stage, a set of couples of potential candidates as coplanar features* $\mathbf{p_s^k}$ *are extracted according to the stereo constraints. They have been tracked between two consecutive images during the third stage. The final stage robustifies coplanar features extraction by imposing the homography constraints to a larger set of couples of view taken in a given horizon time* $T \in [k - k_0, k]$ *with* $k_0 > 2$*.*

The estimation of the stereo-rig egomotion can be splitted in four linked sequential stages:

(1) the segmentation of the road plane in the left and right images by taking account the road geometry and parallel lines of the environment,
(2) the extraction of candidate coplanar features in the segmented region which satisfy the spatial constraints induced by the stereo-rig geometry,
(3) the candidate coplanar features are tracked in the sequence and the false matches corresponding to moving parts in the scene are rejected.
(4) the refinement of the feature selection by using the homography constraints linking multiple views in a bounded horizon time.

3 Segmentation of the road plane

In an urban context, the road is commonly viewed by the on-board cameras as the dark homogeneous region limited by the light road markers, always located on the bottom of the image. Assuming that some of the road markers are visible in the foreground, they can be used to estimate the road direction in order to segment the road area in the images. The identification of the road markers remains nevertheless complex due to presence of kerb-sides, intersection areas or the obstruction by dynamical obstacles. Consequently, using

only a data-based approach for segmenting the road is not robust enough and it will be benefit to introduce some global constraints from a model. Let us assume i) the sides of the road are parallel and can be locally approximated by segments of lines, ii) most of the edges in a caynon-like urban environment are aligned with three main directions: the longitudinal and transversal axis of the road and the vertical axis.

3.1 Extraction of the Dominant Vanishing Point

Under a perspective projection, a set of parallel lines in the 3D scene projects onto the image as a pencil of 2D lines which converge to the same point, called vanishing point. The vanishing point is characteristic of a direction in the scene and it belongs to the plane at the infinity. As the projective homography linking the plane at the infinity and the image plane depends only on the rotations, the vanishing point will be invariant with regard to a translation motion of the camera. Such a property will be useful to detect and track the vanishing point corresponding to the direction of the road axis referred below as the *Dominant Vanishing Point (DVP)*.

Considering our embedded stereo-rig, the variation of the DVP coordinates has two possible origins: the variation of the pan angle when the vehicle is turning and the variation of the tilt angle when the vehicle is accelerating or braking. Assuming the baseline of the stereo-rig is parallel to the road plane and the tilt angle remains small, the model of evolution of the DVP coordinates can be selected as follows: a constant velocity model for abscissas and a stationary model for ordinate. Thanks to a Kalman filtering technique, the estimation and the tracking of the DVP during the vehicle motion can be performed robustly. The observation vector Y is computed from a subset of lines $S = l_0^t \dots l_M^t$ built from the segments extracted into the current image using a Canny operator. To be candidate the lines have to converge in a neighborough of the predicted DVP or to be constituted with segments which have similar characteristics (orientation, gradient, location) than the segments selected at the previous iteration. The observation vector Y is then computed as the image point which minimizes the weighted sum of distances to the subset of selected lines.

$$Y = min_x \sum_{s=1}^{M} (w_s . \mathbf{l_s^t} . \mathbf{x})^2 \tag{1}$$

with w_s a weight that represents the sum of the length of the segments contributing to the line $\mathbf{l_s^t}$.

To be exhaustif on this topic (see [27] for more details), we would like to stress

that the quality of the estimation of the DVP is closely related to the solvability of the equation (1). When all the lines have quite similar orientations in the image (like in Fig. 3), the solution of the equation (1) becomes very instable. In this case, to avoid discontinuities in the DVP estimation due to a lack of constraints, we propagate the prediction of the ordinate provided by the Kalman filter instead of the estimated value.

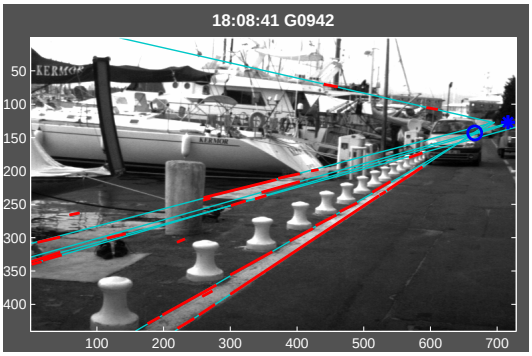


Fig. 3. *Case where the estimation process is not constrained sufficiently.*

The Fig. 4 illustrated the result of the DVP detection and tracking algorithm applied to a stereo video sequence involving an overtaking manoeuver. The vehicle follows a straight path, stops before a parked vehicle, overtakes it then continue its straight motion. The plots represent the variations of the DVP coordinates estimated by the Kalman filter in the left and right images. The smoothness of the chronograms confirm us on the adequation of the models. Complete sequences are available on our web site [1] .

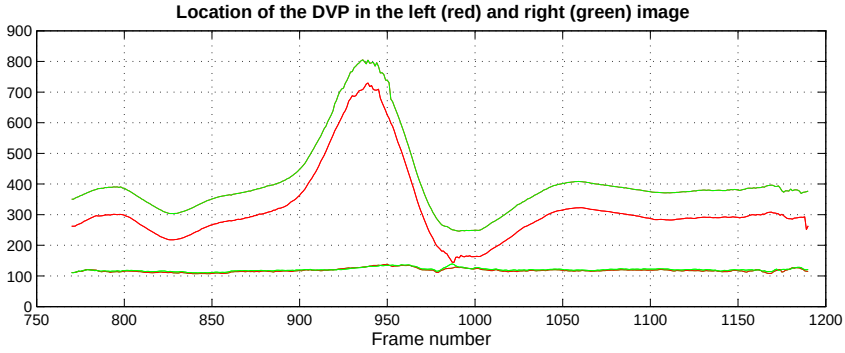


Fig. 4. *Chronogram of the left and right DVP coordinates (abscissas at the top, ordinates at the bottom) during the overtaking maneuver. A DVP abscissa constant corresponds to piece of straight path. The ordinate variations due to the pitching appear limited in comparison of the abscissas ones that justifies the stationary model of evolution. The small variations between the frames [960;990] results from the braking and acceleration phasis during the overtaking maneuver.*

[1] `http://caor.ensmp.fr/~simond`

3.2 Characterization of the traffic lanes.

Most of authors assume that the road limits can be easily detected due to their artificial (road markers) or natural (kerbs, soft shoulders or green strips, sidewalks) contrast with the pavement. The detection of these road markers is intensively used in many ADAS applications or for lateral control in automatic guidance schemes [30,17,2]. We personally suppose that we can track clear or dark stripes with constant width which represent the road or the lanes limits. We aim to use this information in complement to the detection of the DVP point in order to fully characterize the road area at the segmentation level. Unfortunately, in standard traffic conditions, only the lane where the vehicle is moving on, can be expected to be observed by the cameras. The others lanes are almost always occluded by the flow of vehicles and can be perceived from time to time only.

From these reasons, we state the road detection problem as a parameter estimation problem. We assume a region of interest corresponding to the road in the image is defined by the current DVP coordinates and two points for a one lane road (resp. three points for a two lanes road) resulting from the intersection of the median axis of the left and right road markers for a one lane road (resp. the left, middle and right road markers for a two lanes road) with the bottom of the image, like in Fig. 5.

At each abscissas of these points is associated a constant velocity filter and the width of the lane is assumed to be constant. The observation of the filter is computed from the detection in the image of stripes which satisfy a dark-light-dark gradient transition and locally oriented towards the DVP. Moreover, more importance are accorded to the stripes detected close to the bottom of the image (just ahead the vehicle). At the initialization step, a prior knowledge on the width of the road and the number of lanes is provided. In the first couple of images if the observations are sufficient, the tracking is automatically initialized, if not, the stripes have to be pointed out by the user manually. During the tracking process when a stripe jumps outside the limits of the image thanks to the constant width constraint, it remains possible to perform the abscissas prediction of the loss stripe median, according to Fig. 6.

4 Road plane Estimation

The segmentation stage described above, can be viewed as a pre-processing which aims at drastically reducing the search area for the features liable to belong to the road plane. Conversely, due to the presence of dynamic obstacles, all the features inside the region of interest (ROI) are not really lying on the

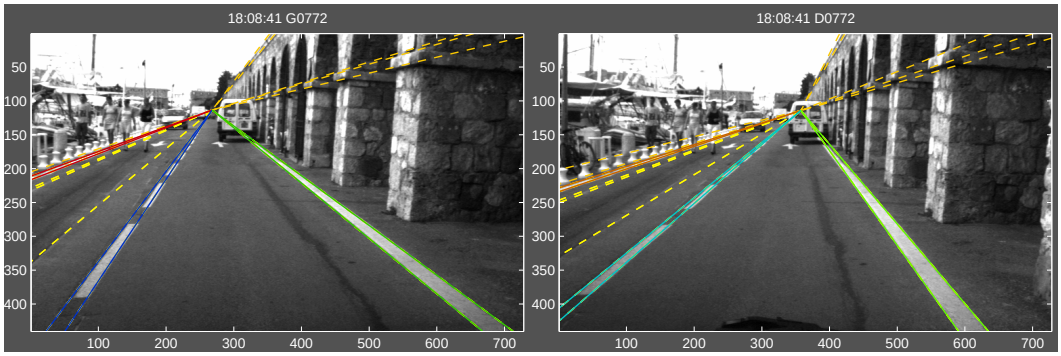


Fig. 5. *Tracked road markers in both images of the stereo-rig in case of a two-lanes road. The dashed lines are others edge-lines which converge to the DVP location.*

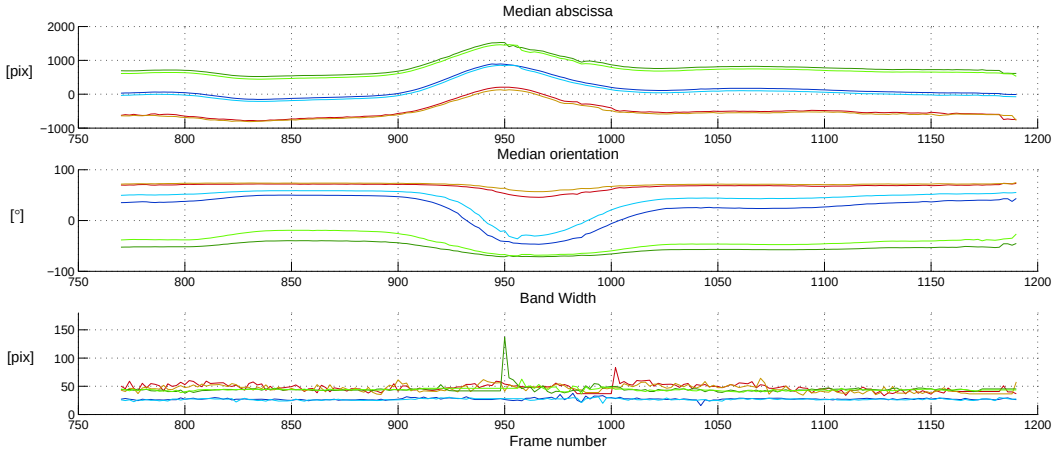


Fig. 6. *Chronogram of the road markers evolution along the sequence. Although some abscissas of the medians of the road markers do not belong to the image limits, the filtering process provides reliable predictions which are very close to the estimated location, even in presence of large rotations (frames [900;1000]). Between the frames [936;952], only the left road marker remains visible in both images. The quality of the estimated abscissas allows the identification of the two others road markers as soon as they go back inside the field of view.*

road plane. In this section, we present a novel method based on the computation of the homographies induced by the road plane in the different views, for classifying the features extracted from the ROI as features lying or not on the road plane. Let us recall a projective homography is a mapping from a projective plane to a projective plane [7] defined up to a scale factor (3×3 matrix but only 8 independent parameters). According that each corresponding points (resp. line) in a couple of images resulting from the projection of a unique point (resp. line) in the scene provides two independent equations of constraint, only four independent points (resp. lines) are required to compute an homography.

The estimation of the camera motion (rotation and translation) $[\mathsf{R_{ba}}, \mathbf{t_{ba}}]$ between the images $\mathcal{I}_a$ and $\mathcal{I}_b$ from the homography $\mathsf{H_{ba}}$, requires the knowledge

of the calibration matrix $\mathsf{K_a}, \mathsf{K_b}$, the unit vector normal $\mathbf{n_a}$ and the distance to the plane d_a, both expressed in the first camera frame:

$$\mathsf{H_{ba}} = \mathsf{K_b} \left[\mathsf{R_{ba}} - \frac{\mathbf{t_{ba}}.\mathbf{n_a^t}}{\mathsf{d_a}} \right] \mathsf{K_a^{-1}} \tag{2}$$

In our particular application, a first idea will be to use the DVP point and the road markers previously extracted for computing the homography. Unfortunately, such a choice is inadequate due to the fact that the projection of parallel lines in the scene yields to linear dependent lines in the images. Others features are needed and we propose to exploit also the Harris' points of interest extracted from the ROI.

4.1 Computing the Homography

Two types of homography are computed at each iteration: the stereo homography $\mathsf{H_{st}^k}$ between the left and right images of the stereo-rig and the homographies $\mathsf{H_l^k}$ (resp. $\mathsf{H_r^k}$) between two consecutive left (resp. right) images. The stereo homography allows the extraction of the coplanar features into the ROI assuming that a reliable prediction is avalaible while the computation of $\mathsf{H_l^k}$ and $\mathsf{H_r^k}$, which are our goals, are performed when the majority of the coplanar features are known in both images. Note that the computation of the homographies $\mathsf{H_l^k}$ and $\mathsf{H_r^k}$ will be useful for rejecting the feature points which belong to the dynamic parts of the scene like other vehicles or pedestrians.

The road pavement appears as a textureless region where the use of SIFT keypoints ([16]) is not well-adapted: the matching of keypoints extracted in this area between two stereo images provides unreliable results. Therefore we use an Harris detector which extract most of the stable feature points (FPs) from the road signs (lines, stripes, arrows, letters, etc.) or from the static and dynamic obstacles present on the road. Unfortunately, only a few FPs extracted from the obstacles do belong to the road plane and the others should be discarded at the processing level.

The accuracy and the stability in the computation of the homography is closely related to the spatial distribution of the FPs into the ROI. To enhance this accuracy, we only keep the most representative FPs in each subregion of a grid applied on the ROI, as in Fig 7. The selection is performed with the Harris score.

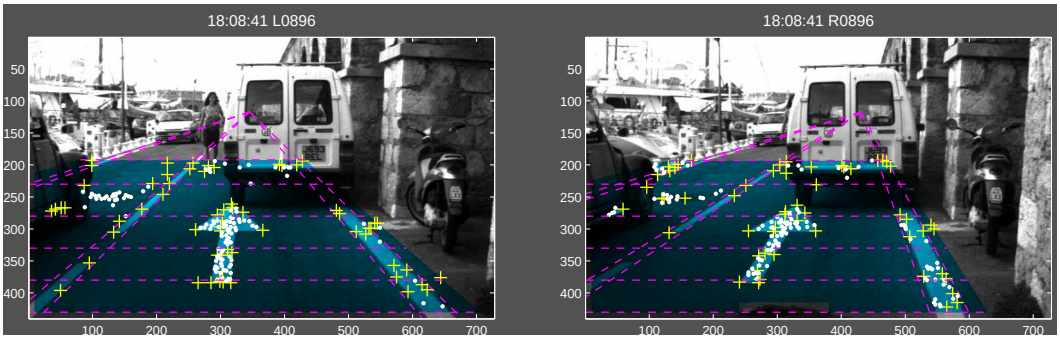


Fig. 7. *The grid applied into the ROI. The road markers provide the majority of the FPs. The road is hence divided into two types of region: those where are located on the road markers and the others. The horizontals split also the ROI with equal height slices that artificially balances the number of FPs detected at the foreground in comparison of FPs detected far from the camera. Some grid areas still remain empty because of the detected FPs have score lower than a critical threshold. Among all the detected FPs '.', only the '+' are selected to compute the current homographies.*

4.2 Homography computed from stereo-vision

Finding correspondences between FPs: The purposed method is based on a variant of the *Iterative Closest Point (ICP)* algorithm [24] that is an alignment algorithm that works in three phases : 1) establish correspondence between pairs of features in the two structures that are to be aligned on proximity, 2) estimate the rigid transformation that best maps the first member onto the second and 3) apply transformation to all features in the first structure. The main difficulty we face is the matching of the two sets of coplanar features which are not sufficiently characterized and uniformly distributed to compute the induced homography.

The Scott Longuet-Higgins ([25]) method consists on the SVD decomposition of a proximity matrix Dist which contains the inter-distance between the two clouds of FPs, weighted by a gaussian.

$$\mathsf{Dist}(i,j) = \exp^{-\left(\frac{dist^2(\mathbf{P_i},\mathbf{P_j})}{2.\sigma^2}\right)} \tag{3}$$

The SVD decomposition verifies $\mathsf{Dist} = \mathsf{U}.\mathsf{D}.\mathsf{V}^{\mathsf{t}}$, the replacement of the diagonal elements of D by an identity matrix $\mathcal{I}$ provides a new proximity matrix $\mathsf{Dist\_svd} = \mathsf{U}.\mathcal{I}.\mathsf{V}^{\mathsf{t}}$ which both respects the exclusion and proximity principles. Each element of $\mathsf{Dist\_svd}$ belongs to the interval [0;1]. Couples of FPs which maximizes simultaneously raw and column of this particular matrix verify the one to one exclusion principle according to the distance criterion.

The algorithm of Pilu ([21]) is dedicated to the matching of stereo images

(small baseline). It performs the last method with purposing an appropriate correspondence strength matrix which integrates both a geometric and an intensity relationship. Parallel to the Dist matrix, the Corr matrix contains the correlation between grey level values of patchs centered on each FP. Due to elements of Corr are normalized between [-1;1], the correspondence strength matrix is:

$$\mathsf{L}(i,j) = \frac{1 + \mathsf{Corr}(i,j)}{2}.\mathsf{Dist}(i,j) \tag{4}$$

All the couples which maximize L_svd, SVD decomposition of L, are not correctly matched. The reliability of the method depends on the threshold we fix to discard couples which can be considered as outliers. Our experiments show that considering a threshold equal to 0.5 provides rarely outliers.

A priori knowledge on the cameras motion: Assuming a rigidity constraint between the two cameras of the stereo-rig, the identification of the road plane and the extraction of coplanar features become easier. The rotation and translation motions $(\mathsf{R_{st}}, \mathbf{t_{st}})$ between the left and right cameras are then fixed and the stereo homography induced by the road plane only depends on the parameters of the plane $(\mathbf{n_{st}^k}, \mathbf{d_{st}^k})$ expressed in the left (right) camera frame. Furthermore, the high frame rate of the video sequence allows us to consider small variations in the parameters of the road plane between two successive image acquisitions.

$$\mathbf{n_{st}^k} \simeq \mathbf{n_{st}^{k-1}} \text{ and } d_{st}^k \simeq d_{st}^{k-1} \tag{5}$$

Consequently, reliable predictions of the stereo homography can be done between two iterations using a simple stationnary model:

$$\hat{\mathsf{H}}_{\mathsf{st}}^{\mathsf{k}} = \mathsf{H}_{\mathsf{st}}^{\mathsf{k-1}} \tag{6}$$

Using this prediction, we can search for potential matches between the left and right images. In practice, the hypothesis of a stationary model does not hold due to the perturbations. Thanks to a recursive estimation algorithm, we refine a posteriori the estimation of the stereo homography which allows us to detect and discard the wrong matches corresponding to FPs which are not really lying on the road plane.

Adaptation of the Pilu method to our case: Assuming the homography prediction, we also search for the correspondences between the two sets of FPs extracted from the left and right current images. A couple of FPs is potentially candidate to the match if and only if it satisfies the crossed re-projection

criterion, like in Fig. 8 :

$$\begin{cases} w_i.w_j.dist(\mathsf{H}_{\mathsf{st}}^{\mathsf{k}}\mathbf{p}_{\mathbf{l_s}}, \mathbf{p}_{\mathbf{r_j}}) < th\_dist \\ w_i.w_j.dist(\mathbf{p}_{\mathbf{l_s}}, (\mathsf{H}_{\mathsf{st}}^{\mathsf{k}})^{-1}\mathbf{p}_{\mathbf{r_j}}) < th\_dist \end{cases} \tag{7}$$

where th_dist is bounded by some pixels and w is a weight which decreases linearly when the ordinate of the selected candidate increases. This ponderation must be introduced to compensate the difference of motion observed in the image according as the point in the scene is more or less close to the vehicle (i.e. the DVP which belongs to the plane at infinity is invariant to the translation motions). All couples of candidates which do not satisfy the distance criterion, are discarded. The second selection rule consists on computing the normalized correlation between grey level of regions $[21 \times 21]\,pixels$ centered on each couple of FPs candidates. Only the couple of FPs satisfying a score greater than $th\_corr = 0.5$ (empirically fixed) are finally labelled as candidates for the matching process. The strength matrix is also sparse that speeds up the SVD computation and reduces the probability of outliers (cf. Fig. 9).

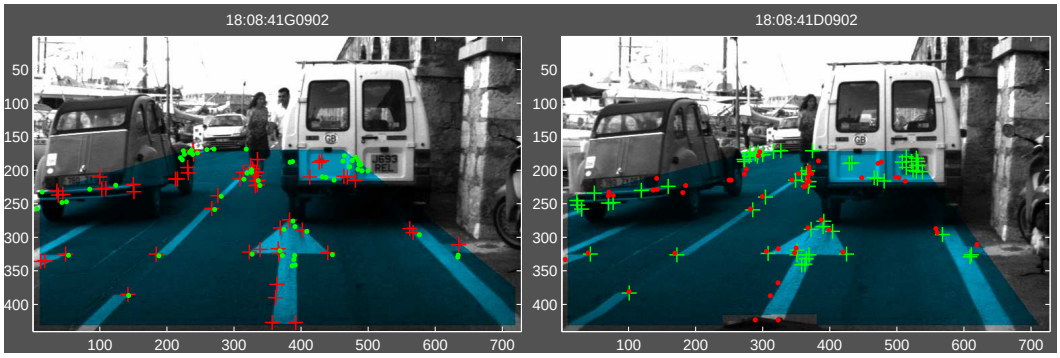


Fig. 8. *Locations of the detected FPs ('+') and their projection ('.') in the other image, computed with the homography prediction* $\hat{\mathsf{H}}_{\mathsf{st}}^{\mathsf{k}}$. *FPs relative to the left and right images are respectively plotted in red and green. Thanks to the uniform spatial distribution of the coplanar features at the last iteration* $(k-1)$*, the high quality of the homography prediction drastically reduces the distance between the coplanar corresponding FPs and improves the relative distance for the others.*

Refining the homography estimation: The quality of the stereo homography computation depends on the spatial distribution of the true coplanar features and on the capability to discard the FPs which are detected on the obstacles over the road plane. The stereo homography computation can hence be resumed as a recursive estimation process which toggles:

(1) the selection of potentially coplanar FPs assuming the current estimation of the homography as described above,
(2) the update of the homography using the couples of FPs matched successfully,

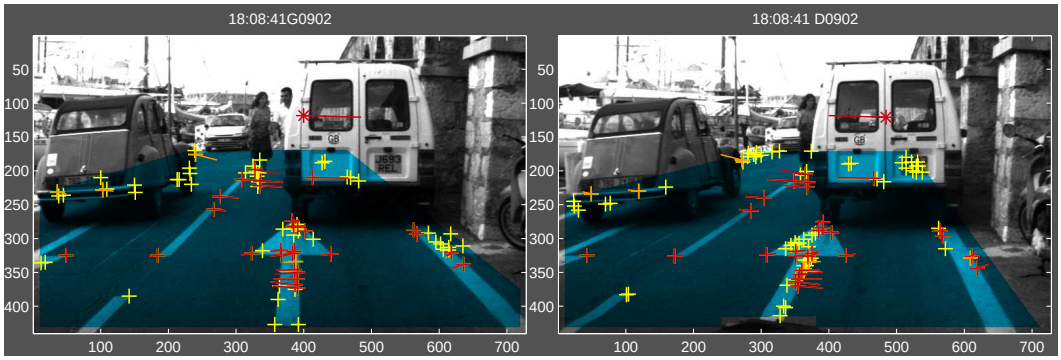
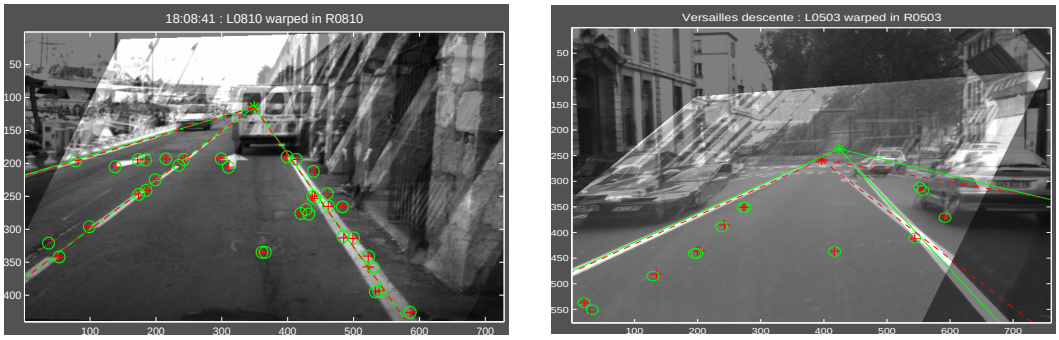


Fig. 9. *Comparaison between the results provided by the implementation of the direct method of Pilu and the advanced one which assume the knowledge of the homography prediction. The detected FPs are shown with yellow '+'. We highlight with red the couples of FPs which are extracted with the advanced Pilu method whereas their relative motion is drawn with a segment. With the direct method, new couples of orange '+' FPs appear which are generally false inliers. In this case, three extra false inliers are detected in complement of the first detected on white vehicle.*

The process stops when the list of selected couples remains stable between two iterations. That is generally the case after two or three iterations (Fig 10(a)). The variance of the distance of re-projection between the couples of FPs is abroad $\sigma = 1\,pixel$, according the weighting introduced sooner. The quality of the homography computation is verified with comparing the orientations of the road markers medians. If the difference is upper than $th\_angle = 10°$, that means the homography computation is not correct due to a non-uniform spatial distribution of the FPs couples, like in Fig 10(b). In such a case, a new initialization of the estimation process is performed.



(a) (b)

Fig. 10. *Warping of the left image into the right one with the stereo homography computation. On the left, the spatial distribution of the selected FPs is uniform, opposite to the case on the right. Homographies are computing with couples of coplanar features: '+' and dashed lines of the left image in correspondence with 'o' and solid lines.*

We would like to stress that the computation of the correlation score between squared patchs centered on FPs lying on a plane which is nearly parallel to the camera axis is not optimal due to the perspective distorsion. Assuming

that a homography prediction is available, a solution consists on computing the warping of the ROI with the homography prediction and then compute the correlation score. The experiments show that when the FPs have not a uniform spatial distribution, the recursive estimation of the homography sometimes diverges.

Initialization step: At the beginning of the sequence, the prediction of the stereo homography is unknown. Assuming that the majority of the detected FPs in the ROI belong to the road plane, we compute a first estimate of the homography using a RANSAC method ([9]). The stereo homography is computed from a set of four couples of FPs (verifying the correlation criterion) selected randomly. The number of random selections is determined assuming a fixed rate of outliers in the sample.

4.3 Homography computed from dynamic vision

The stereo-vision approach provides us with a snapshot where all parts of the environment look like static. The homography computed from stereo-vision gives us an information of the *geometry* of the environment (i.e. partitioning features lying on the same planes). Complementary, the left and right video sequences informs us about the *dynamic* parts of the environment and about the motion of the vehicle. However, only just observing the optical flow field between two successive images, it is not obvious to separate the component due to the camera ego-motion (i.e. induced by static objects) and the component relative to the dynamic objects in the scene. To overpass this difficulty, we propose to track in the video sequence only the FPs satisfying the stereo homography of the road plane.

With regard to the computation of the stereo homography, the difficulty we have to face is the lack on knowledge about the vehicle motion. Consequently, we can not compute a prediction of the homography linking two successive images. As done by several authors [5,20], we assume a forward motion of the vehicle composed by a translation along the longitudinal axis, a rotation around the vertical axis (pan angle) and a rotation around the lateral axis (tilt angle). Thanks to this model, we can define in the current image (k) a bounded region which contains all the potential candidates liable to be matched with a FP in the last image ($k-1$) (cf. Fig.11).

These approximations allow hence the selection of corresponding candidates for each FP selected with the stereo homography in both images. Couples of matched FPs are also extracted with the advanced Pilu algorithm. The homography is then computed with the same method as the stereo homography computation. At this step, the conditions are identical: we benefit from cou-

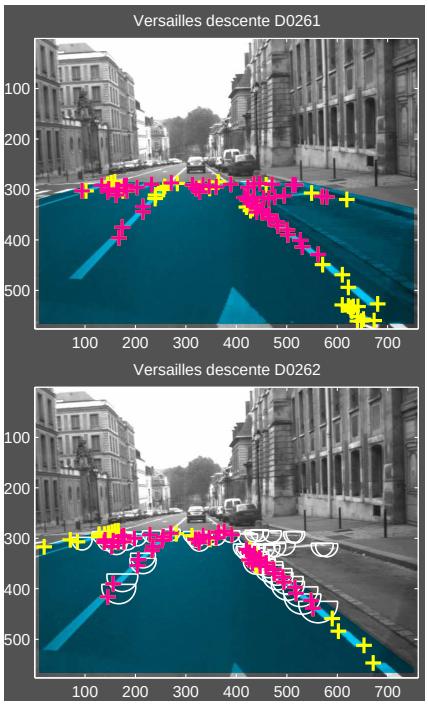


Fig. 11. Prediction areas used in the matching process

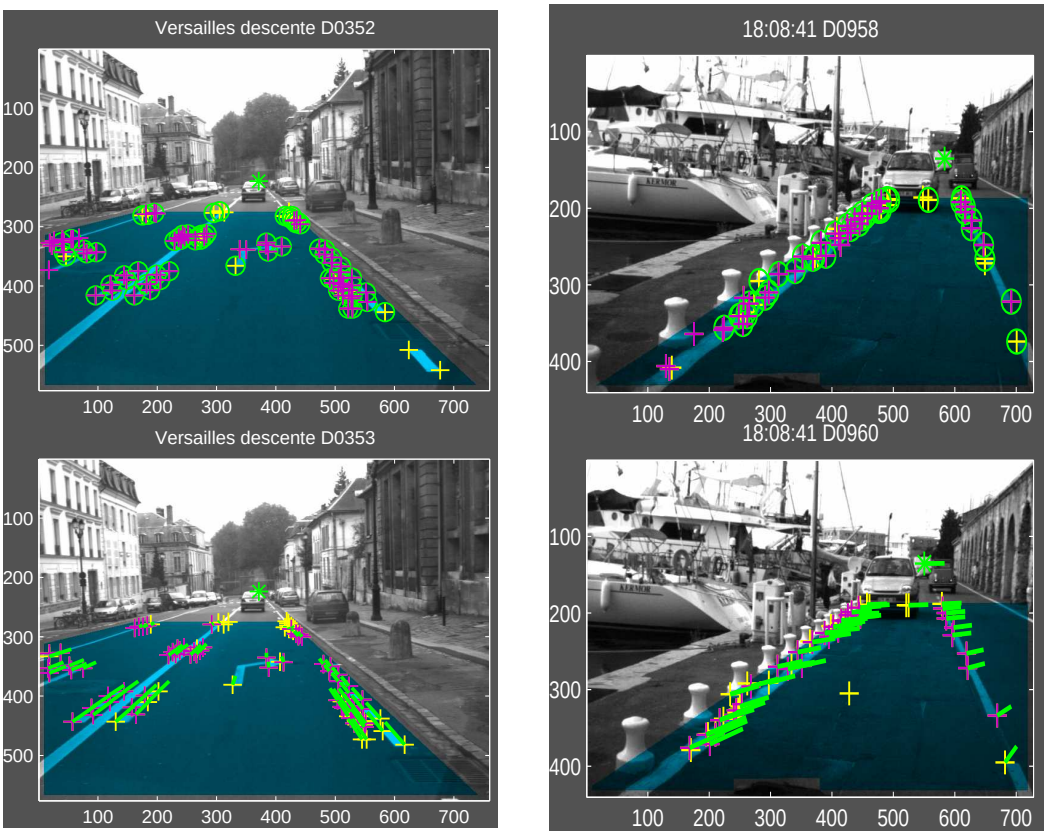


Fig. 12. *Relative motions of the coplanar FPs ('+') according to a straight translation (left) and a dominant rotation motion of the vehicle (right).*

ples of matched FPs, a prediction of homography and characteristics of the road markers. We present in Fig.12 some results of homography computations in basic cases.

5 Integration of spatio-temporal constraints

The stability of the homography computation depends on the coplanar features we achieve to detect. Whatever the type of computed homography may be, some couples of FPs located over the road plane are erroneously selected. We aim at rejecting these outliers for improving the quality of the homographies. The basic idea is to exploit all the constraints linking a larger set of views assuming that the composition rule: $\mathsf{H_{ca}} = \mathsf{H_{cb}}\mathsf{H_{ba}}$ whatever the images $\mathcal{I}_a$, $\mathcal{I}_b$ and $\mathcal{I}_c$ is verified.

5.1 The concept of super-homography

Malis and Cipolla [18] describe an efficient method to impose the constraints between the homographies computed from a sequence of views of a planar structure. Tacking into account multiple views provides a set of constraints between the coordinates of coplanar features in different views. Moreover, it minimizes the effects of errors on the matching step and reduces numerical instability when the motion between two views is not enough significant.

Such a method improves the consistency of the current homographies $(\mathsf{H}^{\mathsf{k}}_{\mathsf{st}}, \mathsf{H}^{\mathsf{k}}_{\mathsf{l}}, \mathsf{H}^{\mathsf{k}}_{\mathsf{r}})$ estimation. Nevertheless, we also face a hard compromise between increasing the distance between two views to reduce the numerical instability and keeping a significant number of matched features to constraint the homographies computations. The method introduces a super-homography matrix $\mathcal{H}^k[3m \times 3m]$ where m is the number of views. This new entity contains all the $[3 \times 3]$ homographies between the $(m(m-1))$ couples of different views. We personally use three stereo couples corresponding to the frames $(k-2)$, $(k-1)$ and (k), the number of images is then $m = 6$. Obviously, solving the constraints imposed by the super homography requires the observation and the tracking of the FPs in the $(k-2)$, $(k-1)$ and (k) couples of images. However, as we show below, the presence of an FP in the whole set of m images is not necessary and only a partial observation in a subset of m will be sufficient in practice.

5.2 The super feature points

At the last stages, the computations of current homographies $(\mathsf{H}^{\mathsf{k}}_{\mathsf{st}}, \mathsf{H}^{\mathsf{k}}_{\mathsf{l}}, \mathsf{H}^{\mathsf{k}}_{\mathsf{r}})$ provide three lists of correspondences between features of the last two couples of stereo-images. The identification and verification of the cross-links between the couples of FPs is processed with fulfilling a table where each column represents a 3D point and the raws the images of the last and current couples. This table allows us to verify the relevance of the current matching steps between the last four images:

$$\mathbf{p}^{\mathbf{k}}_{\mathbf{r_s}} = [\mathsf{H}^{\mathsf{k}}_{\mathsf{st}}.\mathsf{H}^{\mathsf{k}}_{\mathsf{l}}].\mathbf{p}^{\mathbf{k-1}}_{\mathbf{l_s}} = [\mathsf{H}^{\mathsf{k}}_{\mathsf{r}}.\mathsf{H}^{\mathsf{k-1}}_{\mathsf{st}}].\mathbf{p}^{\mathbf{k-1}}_{\mathbf{l_s}} \tag{8}$$

Let us now introduce a new entity: the *super feature point (SFP)* which is a $[3m \times 1]$ vector which contains the homogeneous coordinates in the m images of a point lying on a road plane. Generally, several points are not detected all along the views. In this case, the coordinates of unknown projections will be initialized with zeros and estimated during the computation of the super homography.

5.3 The virtual feature points

As the previous examples shown, the computation of the homography is highly sensitive with respect to the spatial distribution of the 3D points in the plane.

The pencil of road marker medians is certainly the most reliable feature we succeed to track over the video-sequence. Its detection is robust with regard to the local occlusions induced by the presence of obstacles on the road. Thanks to the property, we can define a new type of features, called *virtual feature points (VFPs)*, to constraint the road plane projection in areas where no SFP is detected.

The VFPs are built from the intersections of the pencil of road marker medians with virtual lines lying on the road, defined by couples of coplanar SFPs observed in the m views. Best results are obtained with virtual lines located in the bottom part of the ROI with different orientations. We personally look for nine lines whose orientations are approximatively distribute between $[-45°; 45°]$ with a step of $10°$, like in Fig. 13.

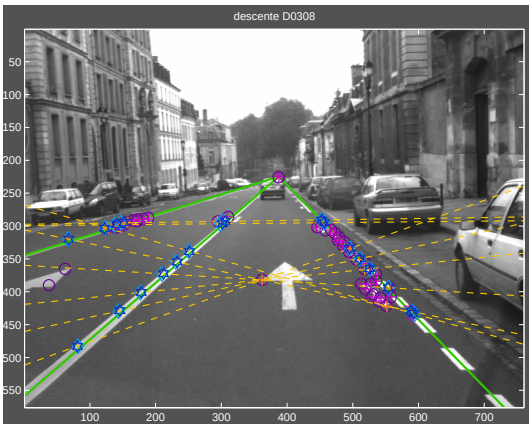


Fig. 13. *Virtual feature points (blue '◇') are intersections of pencil of road markers medians (green solid) and the virtual lines (orange dashed), defined with couples of SFPs, whose projections are known in the m views.*

5.4 The computation of super-homography

The concatenation of the SFPs and VFPs coordinates in the m images composes a global feature matrix ${}^0\mathcal{F}^k$ $[3m \times n_F]$ where $n_F = n_{FP} + n_{VFP}$:

$$ {}^0\mathcal{F}^k = [{}^{\mathbf{0}}\mathcal{P}^{\mathbf{k}}_{\mathbf{1}}, {}^{\mathbf{0}}\mathcal{P}^{\mathbf{k}}_{\mathbf{2}}, ..., {}^{\mathbf{0}}\mathcal{P}^{\mathbf{k}}_{\mathbf{n_{FP}}}, {}^{\mathbf{0}}\mathcal{VP}^{\mathbf{k}}_{\mathbf{1}}, {}^{\mathbf{0}}\mathcal{VP}^{\mathbf{k}}_{\mathbf{2}}, ..., {}^{\mathbf{0}}\mathcal{VP}^{\mathbf{k}}_{\mathbf{n_{VFP}}}] \tag{9} $$

The homographies which form ${}^0\mathcal{H}^k$ are initialized from the composition of homographies estimated at the last and current frames, e.g.:

$$ {}^0\mathcal{H}^k(\mathcal{I}^k_l \mathcal{I}^{k-2}_r) \propto \mathsf{H}^{\mathsf{k}}_{\mathsf{l}}.\mathsf{H}^{\mathsf{k-1}}_{\mathsf{l}}.\mathsf{H}^{\mathsf{k-2}}_{\mathsf{st}} \tag{10} $$

At this point, SFPs and VFPs are both vectors of projections coordinates observed in the m images. The first estimation of each column of ${}^1\mathcal{F}^k$ is also computed according to some projection coordinates are unknown:

$$ {}^{\mathbf{1}}\mathcal{P}^{\mathbf{k}}_{\mathbf{s}} \propto \frac{1}{m^*} {}^0\mathcal{H}^k.{}^{\mathbf{0}}\mathcal{P}^{\mathbf{k}}_{\mathbf{s}} \tag{11} $$

where $m^* \leq m$ represents the number of known projections of the s^{th} SFP with $s \in [1; n_F]$.

When the composition rule is verifyied for all homographies, $rank(\mathcal{H}^k) = 3$ whatever $m \geq 3$. Due to all the numerical errors, $rank({}^0\mathcal{H}^k) > 3$. Therefore an iterative estimation is required which toggles until $q \geq 2$ with the estimation of SFPs coordinates according to the last super-homography computation and the computation of all the homographies induced by SFPs projections in all

images:

$$\begin{cases} {}^{q}\mathcal{F}^{k} \propto \frac{1}{m}\, {}^{q}\mathcal{H}^{k-1}.{}^{(q-1)}\mathcal{F}^{k} \\ {}^{(q-1)}\mathcal{F}^{k}(\mathbf{M_r},:) \propto {}^{q}\mathcal{H}^{k}(\mathbf{M_r},\mathbf{M_c}).{}^{(q-1)}\mathcal{F}^{k}(\mathbf{M_c},:) \end{cases} \tag{12}$$

where $\{\mathbf{M_r}, \mathbf{M_c}\}$ respectively represent numbers of raws and numbers of columns relative to the images $\mathcal{I}_{m_r}$ and $\mathcal{I}_{m_c}$ for the super-homography with $\{m_r, mc\} \in [1; m]$ and $m_r \neq m_c$.

The process stops when $rank({}^{q}\mathcal{H}^{k}) = 3$, that is generally the case after $q = 3$ iterations unless the system was ill-conditioned due to errors on feature matching or lack of constraints with the features spatial distribution. The ideal sub-pixellic homogeneous coordinates of the features in the current stereo images can be extracted from ${}^{q}\mathcal{F}^{k}$ while the current homographies can be extracted from the final estimation of ${}^{q}\mathcal{H}^{k}$:

$$\mathcal{H}^{k} = \begin{bmatrix} \mathsf{I}_3 & (\mathsf{H}_{\mathsf{st}}^{\mathsf{k-2}})^{-1} & \cdots & \cdots & \cdots & \cdots \\ \mathsf{H}_{\mathsf{st}}^{\mathsf{k-2}} & \mathsf{I}_3 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \mathsf{I}_3 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \mathsf{I}_3 & \vdots & \vdots \\ \cdots & \cdots & \mathsf{H}_{\mathsf{l}}^{\mathsf{k}} & \cdots & \mathsf{I}_3 & (\mathsf{H}_{\mathsf{st}}^{\mathsf{k}})^{-1} \\ \cdots & \cdots & \cdots & \mathsf{H}_{\mathsf{r}}^{\mathsf{k}} & \mathsf{H}_{\mathsf{st}}^{\mathsf{k}} & \mathsf{I}_3 \end{bmatrix} \text{ with } \mathcal{P}_{\mathbf{s}}^{\mathbf{k}} = \begin{vmatrix} \mathbf{p}_{\mathbf{l_s}}^{\mathbf{k-2}} \\ \mathbf{p}_{\mathbf{r_s}}^{\mathbf{k-2}} \\ \vdots \\ \vdots \\ \mathbf{p}_{\mathbf{l_s}}^{\mathbf{k}} \\ \mathbf{p}_{\mathbf{r_s}}^{\mathbf{k}} \end{vmatrix} \tag{13}$$

Note that the upper matrix of $\mathcal{H}^{k}$ is the inverse transpose of the lower part that makes the ${}^{0}\mathcal{H}^{k}$ computation easiest.

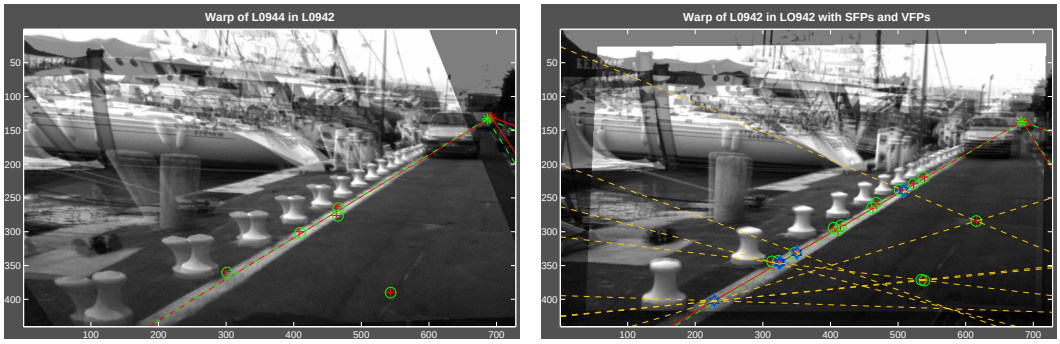


Fig. 14. *Influence of the feature spatial distribution to compute homographies. The hybrid images are composed with current image (green features) warped into the last image (red features) according to the induced homography* $\mathsf{H}_{\mathsf{l}}^{\mathsf{k}}$ *(on left) and the corresponding homography extracted from the super-homography* $\mathcal{H}^{k}$ *(on right). Orange dashed lines are virtual lines that generate the SFPs in current (light blue '◇') and last (dark blue '⋆') image.*

The Fig. 14 shows the utility of the VFPs computation when the FPs spatial distribution is not uniform in the ROI or when obstacles obstruct the ROI clear view. Initially, four of the five matched FPs were aligned with the road marker, the induced linear system was ill-conditioned. The introduction of the SFPs, light blue '$\diamond$' and dark blue '$\star$' respectively in current and past image, drastically improves the quality of estimation: pencil of VLs correctly superimpose and new couples of FPs are identified. Among them two are not aligned with the left road marker.

The pseudo-code of the super-homography computation is:

(1) q*=0,*
(2) *fullfilling the table to verify the relevance between the couple of corresponding FPs in the last and current couples of images,*
(3) *creation of* ${}^0\mathcal{F}^k$ *from* ${}^0\mathcal{F}^{k-1}$ *and colums of the table,*
(4) *selection of the SFPs* ${}^0\mathcal{P}^k$ *whose projections are known in the* m *views to complete* ${}^0\mathcal{F}^k$ *with the VFPs* $\mathbf{{}^0\mathcal{VP}^k}$*,*
(5) *initialization of* ${}^0\mathcal{H}^k$ *with* $\mathcal{H}^{k-1}$ *and* $\mathsf{H}^{\mathsf{k}}_{\mathsf{st}}, \mathsf{H}^{\mathsf{k}}_{\mathsf{g}}, \mathsf{H}^{\mathsf{k}}_{\mathsf{d}}$*,*
(6) **while** $rank({}^q\mathcal{H}^k) > 3$*,*
  (a) $q = q+1$*,*
  (b) *computation of features coordinates* $\mathsf{{}^qF^k}$ *from* ${}^q\mathcal{F}^{k-1}$ *and* ${}^q\mathcal{H}^{k-1}$*,*
  (c) *estimation of each* $[3 \times 3]$ *matrix of* ${}^q\mathcal{H}^k$ *from* ${}^q\mathcal{F}^k$*,*
(7) **end while**
(8) *searching the SFPs whose the estimated coordinates* $\mathbf{{}^q\mathcal{P}^k_s}$ *are far from the initial coordinates* $\mathbf{{}^0\mathcal{P}^k_s}$ *in at least one of the* m *images,*
(9) **if** *false matchs appear,*
  (a) *reset wrong projections in* $\mathsf{{}^0F^k}$*,*
  (b) *new* $\mathcal{H}^k$ *estimation from the modified* ${}^0\mathcal{P}^k_s$ *and the estimated* ${}^q\mathcal{H}^k$*,*
(10) **end if**
(11) *fullfilling the features structure*
  (a) *updating the SFPs fields with their projections coordinates in the current stereo images,*
  (b) *creation of new SFPs elements whose projections are known in at least* $m/2$ *images with a relationship between 2 stereo images.*

5.5 Homography decomposition

Assuming that the intrinsic parameters of the camera(s) are known, upper triangular matrix, the rotation and translation motion of the camera which induces the homography can be retrieved from (2). Due to the homography depends on 8 parameters, its decomposition requires an a priori knowledge of the normal to plane to distinguish the right solution ([8]) among the two transformation couples which rely on **n** and (**-n**).

6 Experimental Result

6.1 Trajectography of the vehicle

Assuming the intrinsic parameters of the cameras are known, the introduction of the super-homography and the VFPs allow the vehicle trajectory calculation along the difficult sequence of Antibes. The Fig. 15 shows the estimated motions between two consecutive left images. The homographies are extracted from the super-homographies The three chronograms successively represent the translation motion **t**, the rotation motion R and the normalized coordinates of the normal to plane **n** (the vertical). The Fig. 16 show the 2D plots of the trajectory along a distance about 200 meters.

The estimated trajectory is conform to the ground truth: a straight path with an overtaking maneuver of a parked car (complete sequences are available on our web site [2]). As shown in the figures, the vehicle motion is mainly a translation along the z-axis with a rotation around the y-axis. The vehicle slows down, stops then accelerates to finally run with a constant speed. The discontinuities observed between frames $[936; 952]$ are due to the reduced view of the road region when the vehicle filters for the overtaking maneuver: the detected FPs are mainly aligned with the left road marker. The jittering effect on the computed normal to plane is relative to the numerical noise introduced by the short motions of features when the vehicle runs slowly. We observe that this jitter is more reduced with the decomposition of the stereo homography.

6.2 Reconstruction of the road plane

As a straightforward result of the homography-based method, we are able to reconstruct a 3D model of the road plane in the scene and warp on it the images

[2] `http://caor.ensmp.fr/~simond`

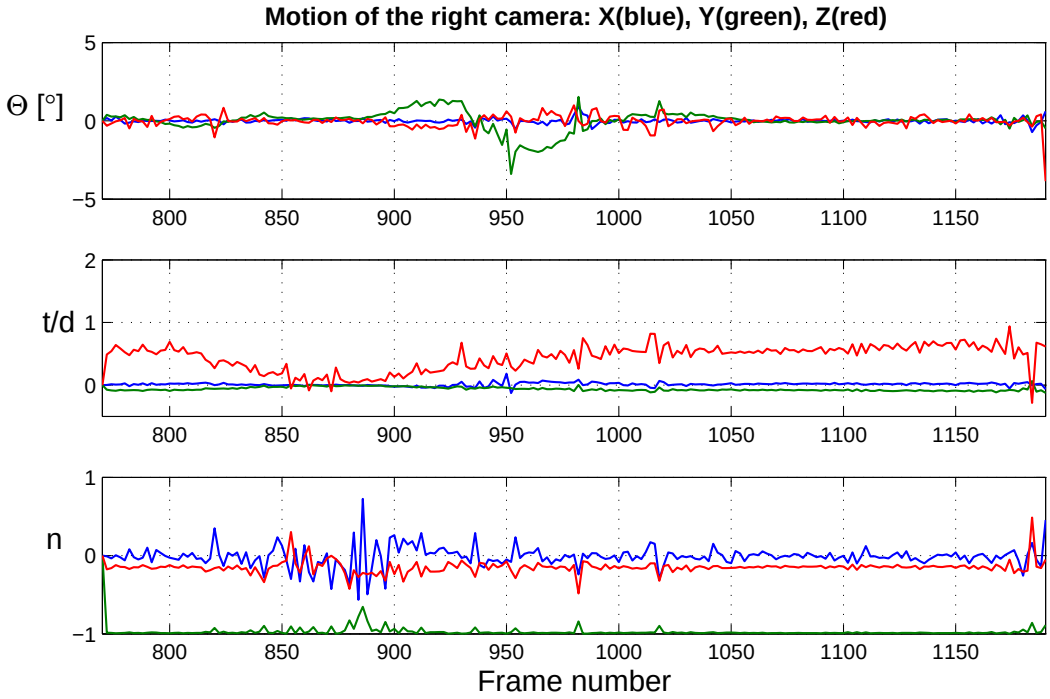


Fig. 15. *Decomposition of homographies between two consecutive right images: rotation motions (top), translation motions (middle), orientations of the normal to plane (bottom).*

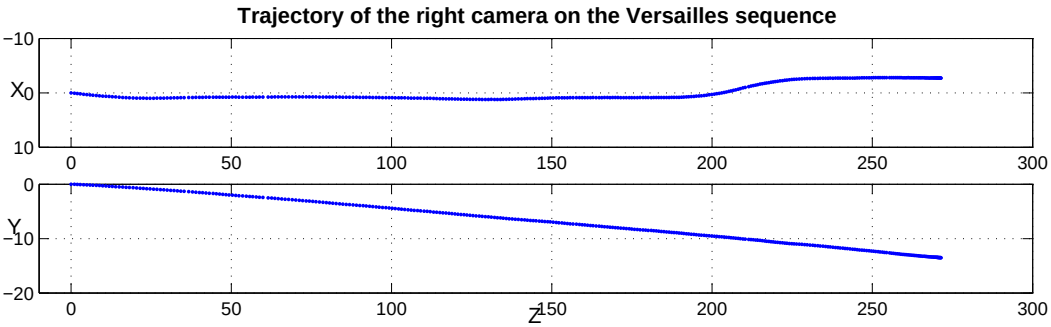


Fig. 16. *2D trajectory of the vehicule motion along the Versailles sequence. The original orientation of the right camera are estimated to* $[2; -6.5; 0]^\circ$ *to obtain as estimation close to the reality.*

taken during the motion (see Fig. 17). We expect that such a representation could have great uses for the urban policy of the cities: the pavement service, cadastre, real-time navigation systems providers. Direct measures are indeed affordable in hybrid image, that offers new solutions for these users. Each frame is also represented in the first view of the sequence with considering the composition rule of homographies between the first and the current frame.

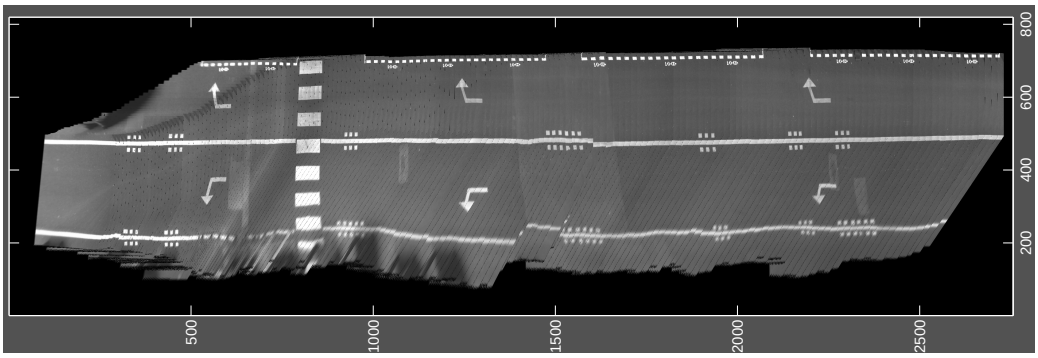


Fig. 17. *Reconstruction of the road plane in a Versailles street. The sequence represents a forward motion of the vehicle along more than* 200 *m with a change of lane at the end. The bird eye view of the road made of the warping of 170 images of the ROI segmented in the right images of the stereo-rig. The discontinuity of the road at the middle of the image is due to a lack of texture in the preceding road area. The extracted SFPs do not constraint sufficiently the super-homography which is finally incorrect.*

6.3 2.5D modeling of the environment

In the specific context of urban environment, a 2.5D modeling can be performed according to the frontages of buildings are mainly composed with vertical planes. The frontages are structured areas where several corners and edges can be extracted. The method developed on the road plane can hence be easily transposed to the vertical planes. Opposite to the FPs lying on the road plane, the FPs extracted on the frontages have a radial motion with a dominant horizontal motion towards the vertical boundaries of images.

We show in Fig. 18 the result of super-homography computation applied on the three main planes of the scene. The vertical planes are first segmented according to the hallway model applied on the environment after the detection of the road area. The verification of the SFPs planarity allows the identification of new regions in images which are not projections of the segmented planes: they are assumed as obstacles. The obstacle detection is the required task to perform the robust navigation of a vehicle. In a piecewise planar scene, this method provides rapidly reliable results.

The localization of the vehicle assuming the super-homography computation from the frontages is nevertheless limited because of the obstacles which obstruct the clear view of the frontages and due to the distance between frontages and cameras which induce reduced magnitude motion of the FPs lying on the frontages. Due to the relief, the homographies induced by the frontages are also less accurate than the homography induced by the road: they could be used as predictions on the vehicle motion when the road plane is occluded.

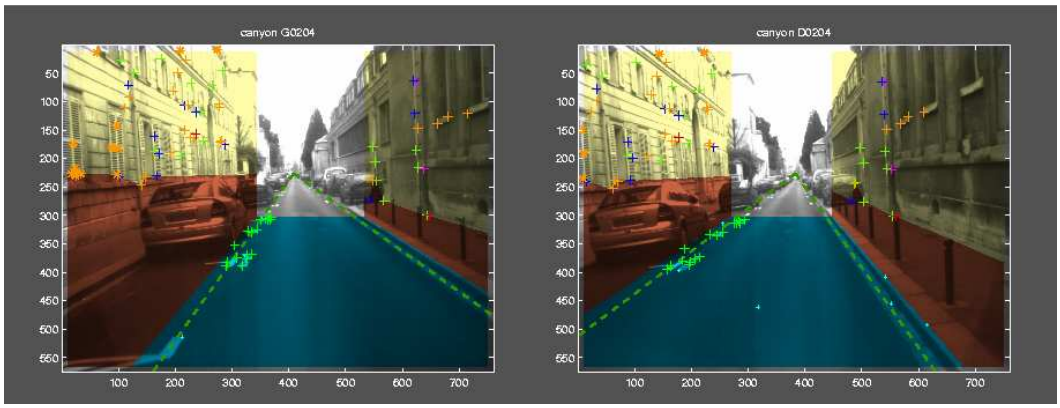


Fig. 18. *According to the hallway model, the modeling of urban canyons with three main planes allows a robust segmentation of the static scene. FPs extracted on areas which are not projections of these planes have to be viewed as obstacles.*

7 Conclusion and perspectives

We detail in this article what sort of informations can be extracted from an on-board stereo-rig. Assuming that urban environments can be assumed as a piecewise planar environment, we present promising results which deals with the autonomous guidance of a vehicle in the urban conditions. The main goal of our work is a vision-based localization method which estimate the vehicle motion according to the tracking of the static scene. Highly urbanized environments can be generally modeled with a hallway model and have structured roads where the lanes are distinguished with road markers. In dense traffic conditions, the clear view of the static scene is limited. We nevertheless develop a method which segment the road area in images to extract coplanar features (lines and points) according to the stereo-vision constraints. The vehicle motion is then estimated with computing the homography induced by features lying on the road plane between two frames.

Dynamic and static obstacles reduce the clear view of the structured static scene. They provide furthermore features which corrupt the matching stage of coplanar features. The homographies computation are also improved by verifying the features coplanarity in multiple views. The super-homography allows it in a single computation and improve the reliability of the coplanar features extraction. The method can be easily transposed to each plane which structured the static scene.

Two material drawbacks limit for the moment the method. First, the field of view of the road is too restricted when the vehicle turns. The use of wide field cameras or the introduction of a third camera certainly allows a widening of the field of view that can maintain an optimal view of the road area in images. Second, the major difficulty we face is to do not access to a prediction of the vehicle motion between two consecutive frames. We plan to synchronize

the frames acquisition with an odometer pulse that has the great advantages to allow a reduction of the frame-rate when the vehicle is not running and otherwise to bundle the relative motion of the static environment between images.

In case of post-processing applications, an accurate segmentation of the road plane can be performed assuming homography computations with direct methods ([28]). Direct methods are indeed well-adapted to textureless area but requires rectified images of the camera and a calibration stage which can be provide by our method. Otherwise, the homography parametrization developed by [3] should bundle the variations of the homography parameters and avoid discontinuities in estimated motions. Auto-calibration of the stereo-rig can be performed assuming that the parallelism and orthogonality of some planes which structure the urban environments. [26] exploit the interest to perform an on-board camera calibration when stairs or pedestrian crossing areas are observed.

References

[1] M.E. Antone and S. Teller. Automatic recovery of relative camera rotations for urban scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'00)*, pages 282–289, Head Island, SC, USA, June 13-15 2000.

[2] R. Aufrère, R. Chapuis, and F. Chausse. A model-driven approach for real-time road recognition. *Machine Vision and Applications*, 2(13):95–107, 2001.

[3] S. Benhimane and E. Malis. Mise en correspondance d'images à différentes résolutions à l'aide d'invariants aux paramètres intrinsèques. In *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA'04)*, Toulouse, France, Jan. 27-30 2004.

[4] F. Berefelt, B. Boberg, J. Nygrds, P. Strmbck, and S.-L. Wirkander. Collaborative gps/ins navigation in urban environment. In *ION National Technical Meeting*.

[5] A. Branca, E Stella, and A Distante. Mobile robot navigation using egomotion estimates. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'97)*, Grenoble, France, Sep. 8-12 1997.

[6] T. Chen. Development of a vision-based positioning system for high density area. In *Asian Conference on Remote Sensing (ACRS'99)*, Hong Kong, China, Nov 22-25 1999.

[7] O. Faugeras. *Three-Dimensional Computer Vision : a Geometric Viewpoint*. MIT Press, 1993.

[8] Olivier Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. Technical report, INRIA, 1998.

[9] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *ACM*, pages 381–395, June 1981.

[10] A. Georgiev and P.K. Allen. Localization methods for a mobile robot in urban environments. In *IEEE Transactions on Robotics and Automation (ICRA'04)*, 2004.

[11] J.J. Guerrero and C. Sagues. *Uncalibrated vision-based on lines for robot navigation*, volume 6, pages 759–777. Elsevier, mechatronics 11 edition, Sept 2001.

[12] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univerity Press, 2003.

[13] M. Kais, S. Morin, A. de la Fortelle, and C. Laugier. Geometrical model to drive vision systems with error propagation. In *8th International Conference on Control, Automation, Robotics and Vision (ICARCV'04)*, Kunming, China, Dec. 3-9 2004.

[14] H. Katsura, J. Miura, M. Hild, and Y. Shirai. A view-based outdoor navigation using object recognition robust to changes of weather and seasons. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'03)*, pages 2974–2979, Las Vegas, Nev., USA, Oct. 27-31 2003.

[15] Qifa Ke and Takeo Kanade. Transforming camera geometry to a virtual downward-looking camera: Robust ego-motion estimation and ground-layer detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, pages 390–7, Madison, USA, June 2003.

[16] David G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*.

[17] Jianye Lu, Ming Yang, Hong Wang, and Bo Zhang. Vision-based real-time road detection in urban traffic. In *The International Society for Optical Engineering (SPIE'02)*, pages 75–82, San-Jose, CA, USA, 2002.

[18] E. Malis and R. Cipolla. Multi-view constraints between collineations: application to self-calibration from unknown planar structures. In *European Conference on Computer Vision (ECCV'00)*, volume 2, pages 610–624, Dublin, EIRE, June 2000.

[19] M. Okutomi, K. Nakano, J. Maruyama, and T. Hara. Robust estimation of planar regions for visual navigation using sequential stereo images. In *IEEE International Conference on Robotics and Automation (ICRA'02)*, pages 3321–3327, Washington DC, USA, May 2002.

[20] N. Pears and L. Bojian. Ground plane segmentation for mobile robot visual navigation. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'01)*, pages 1513–1518, Maui, HI, USA, Oct. 29-Nov 3 2001.

[21] M. Pilu. A direct method for stereo correspondence based on singular value decomposition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 261–266, San Juan, Puerto Rico, June 17-19 1997.

[22] E Royer, M Lhuilier, M. Dhome, and T. Chateau. Localization in urban environments: monocular vision compared to a differential gps sensor. In *15th British Machine Vision Conference (BMVC'04)*, London, U.K., Sept. 7-9 2004.

[23] E Royer, M Lhuilier, M. Dhome, and T. Chateau. Towards an alternative gps sensor in dense urban environment from visual memory. In *15th British Machine Vision Conference (BMVC'04)*, London, U.K., Sept. 7-9 2004.

[24] Szymon Rusinkiewicz. Rigid-body alignment. In *IEEE International Conference on Computer Vision (ICCV'05)*.

[25] G. Scott and H. Longuet-Higgins. An algorithm for associating the features of two patterns. In *Royal Society of London*, volume B244, pages 21–26, 1991.

[26] S. Se and M. Brady. Road feature detection and estimation. *Machine Vision and Applications*, 14(3):157–165, 2003.

[27] N. Simond and P. Rives. Trajectography of an uncalibrated stereo-rig in urban environments. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'04)*, Senda, Japan, Sept. 2004.

[28] G.P. Stein, O. Mano, and A. Shashua. A robust method for computing vehicle ego-motion. In *IEEE Intelligent Vehicle Symposium (IV'00)*, pages 362–368, Piscataway, NJ, USA, Oct. 3-5 2000.

[29] A. Talukder, S. Goldberg, L. Matthies, and A. Ansar. Real-time detection of moving objects in a dynamic scene from moving robotic vehicles. In *IEEE RSJ/International conference on Intelligent Robot and System (IROS'03)*, pages 1308–1313, Las Vegas, Nev., USA, Oct. 27-31 2003.

[30] C.J. Taylor, J. Malik, and J. Weber. A real-time approach to stereopsis and lane-finding. In *IEEE Intelligent Vehicles Symposium (IV'96)*, pages 207–212, New-York, USA, 1996.

[31] W. van der Mark, D. Fontijne, L. Dorst, and F.C.A. Groen. Vehicle ego-motion estimation with geometric algebra. In *IEEE Intelligent Vehicle Symposium (IV'02)*, pages 58–63, Versailles, France, June 17-21 2002.

[32] C.-C. Wang. *Simultaneous Localization, Mapping and Moving Object Tracking.* PhD thesis, Robotics Institute, Carnegie Mellon University, 2004.

[33] C.-C. Wang, C. Thorpe, and S. Thrun. Online simultaneous localization and mapping with detection and tracking of moving objects: Theory and results from a ground vehicle in crowded urban areas. In *IEEE International Conference on Robotics and Automation (ICRA'03)*.