



Reconstruction of the road plane with an embedded stereo-rig in urban environments

Nicolas Simond

► To cite this version:

Nicolas Simond. Reconstruction of the road plane with an embedded stereo-rig in urban environments. IEEE Intelligent Vehicle Symposium (IV'06), Jun 2006, Tokyo, Japan, pp.70-75. inria-00091220

HAL Id: inria-00091220

<https://inria.hal.science/inria-00091220>

Submitted on 25 Sep 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Reconstruction of the road plane with an embedded stereo-rig in urban environments

Nicolas SIMOND

Centre de Robotique, Ecole Nationale Supérieure des Mines de Paris, 60 Bd Saint-Michel, 75 272 Paris Cedex 06, France

Email: Nicolas.Simond@ensmp.fr

Abstract—We present in this article an original method to reconstruct the road in the specific context of urban environment thanks to the data provided by an uncalibrated stereo-vision system. The method consists on extracting then tracking features (points, lines) from the road and estimate the homography induced by the plane between two poses. 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. Experimental issues from real data are presented and discussed.

I. INTRODUCTION

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 environment, this process is enhanced using a priori knowledge from maps thanks to a map matching algorithm which constraints the trajectory of the vehicle to belong to the road. Practically, in a dense urban environment, the reliability of the GPS data and its availability 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 multi-paths 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.

Basically, on-board vision was implemented to improve the robustness of the navigation task by analyzing the evolution of the environment. Vision has 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 egomotion of the camera(s) assuming a prior knowledge on the environment:

- optic flow ([15]) is used in case of unstructured environment or textureless environment,
- direct method ([14], [7]) provides reliable results assuming a calibrated camera and a motion model reduced to the three main parameters,
- homography ([9]) to detect planes in the environment and compute then the relative motion of the camera.

Thanks to the fact that urban environment are structured, the localization of a vehicle is affordable assuming a trans-

position of visual methods dedicated to indoor environment. The relative camera orientation can be computed from the location of vanishing points ([1]). 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 ([4]). 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 ([2], [5]). Another way consists on the update of the current view with a set of images, recorded and geo-referenced during a calibration phase ([6], [11]).

The paper is organized as follows. In the next section, we describe an overview of the method and the assumptions we make. The third section details the main part of the method: segmentation of the road plane, extraction of coplanar features in stereo images then between two consecutive images to finally explain how the introduction of super-homography improves the robustness of the last stages. In the fourth part, we present an example of reconstruction of the road plane. We conclude this article in the last section by proposing perspectives.

II. ASSUMPTIONS AND OUTLINES

A. Road model

This article deals with the specific context of urban environment where the field of view is limited by the dynamical obstacles (vehicles, pedestrians) and the structurant elements of the static environment. We hence assume that the clear field of view do not exceed ten meters in front of the vehicle. The main assumptions we make are common to most of authors who deal with navigation in structured environments:

- the road is locally planar, has parallel borders and may contain painted road markers which have normalized width,
- the camera is modeled with a pinhole model,
- the frame rate of the video sequence induces small variations on the image characteristics,
- the vehicle do not exceed the speed limit in urban environment (abroad 50 km/h in European countries).

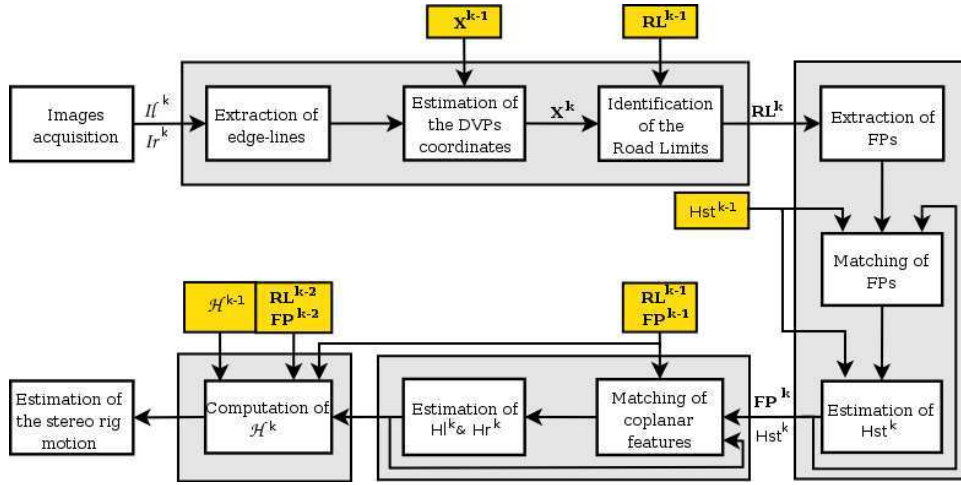


Fig. 1. Schematic diagram of the method. The first stage corresponds to the segmentation of the road plane in images \mathcal{I}_l^k and \mathcal{I}_r^k . In the second stage, couples of the coplanar features \mathbf{p}_s^k are extracted according to the stereo constraints. They have been tracked between two consecutive images during the third stage. The final stage provide robust coplanar features with verifying the coplanarity of the current features in several couples of views.

B. Outlines

Assuming that we can track features lying on the plane, it is then possible to estimate the relative pose and orientation of a camera with considering the homography which characterizes the projection of each 3D plane in the retinal plane of the camera. The relative motion of the camera can thus be deduced from the relative motion of the road plane in images. The estimation of the stereo-rig motion between two poses can be divided in four linked sequential stages according to the synoptic in Fig.1.

C. Notations

The relative motion of a plane π induces an homography H_{ba} between its projection in couples of images \mathcal{I}_a and \mathcal{I}_b . Only the coplanar features of both images are matched according to the projective equalities:

$$\mathbf{p}_b \propto H_{ba} \mathbf{p}_a \quad \text{and} \quad \mathbf{l}_b \propto H_{ba}^{-t} \mathbf{l}_a \quad (1)$$

where \mathbf{p}_γ and \mathbf{l}_γ are respectively the projections of a 3D point and a 3D line, lying on the π plane, observed in the image \mathcal{I}_γ . The recovery of the transformation $[R_{ba}, \mathbf{t}_{ba}]$ between the two camera(s) poses from the homography H_{ba} , requires the knowledge of the upper triangular matrix K_a, K_b containing the intrinsic parameters, the unit vector normal \mathbf{n}_a and the distance to the plane d_a , both expressed in the first camera framework:

$$H_{ba} = K_b \left[R_{ba} - \frac{\mathbf{t}_{ba} \cdot \mathbf{n}_a^t}{d_a} \right] K_a^{-1} \quad (2)$$

In the Fig.2, the poses of the stereo-rig is shown in the last $(k-1)$ and current (k) iterations. Homographies induced by the π plane between the stereo images are noted H_{st} , whereas homographies between consecutive left images (resp. right) are noted H_l (resp. H_r). Given that the stereo-rig is fixed behind

the windshield, the motion between the two cameras can be assumed as rigid:

$$R_{st}^k \equiv R_{st}^{k-1} \quad \text{and} \quad \mathbf{t}_{st}^k \equiv \mathbf{t}_{st}^{k-1} \quad (3)$$

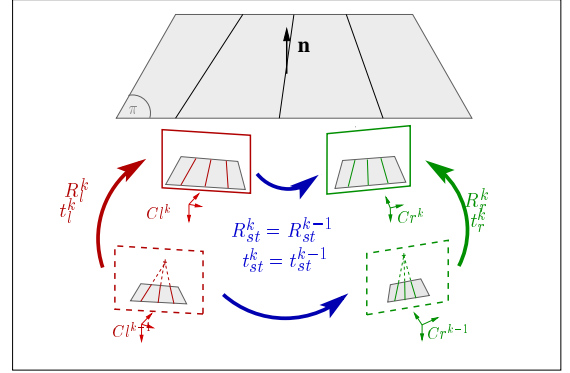


Fig. 2. Main motions between two poses of the stereo-rig. The method distinguish three types of relative motion between couples of views: the stereo motion ($_{st}$), the camera motion ($_l$ or $_r$) and all others (not shown) which are compositions of the first four.

III. METHODOLOGY

A. Segmentation of the road plane

We detailed in a precedent article [13] how to extract the road plane boundaries in images from edges provided by a Canny edge detector. We introduced the Dominant Vanishing Point which is the intersection between all the 3D lines aligned with the road direction and the retinal plane. According to the relative motion of the plane is a priori reduced between two images and a majority of edges extracted at the foreground are aligned with the road direction, we use a Kalman filtering process to obtain a prediction of the DVP coordinates in the next image considering the current ones. Unfortunately, the

edges are not reliable therefore we have to merge first them to form the projections in images of the 3D lines.

We show in Fig. 3, the result of the DVP location in both images along the Antibes sequence. The vehicle follows a straighted path, stops before a parked vehicle, overtakes it then continue its straighted motion. The curbs represent the coordinates of the DVP measured, predicted and estimated by the Kalman filter in both images. The smoothness of the chronograms allow us to confirm our assumptions and models. Animations on the sequences are available on our web site¹.

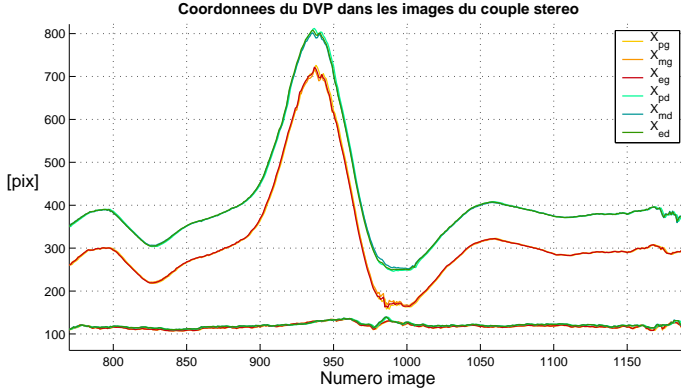


Fig. 3. Chronogram of the left (red) and right (green) DVP coordinates (abscissas at the top, ordinates at the bottom) along the Antibes sequence. The ordinate variations appear limited in comparison of abscissas ones that justifies the invariant model of evolution.

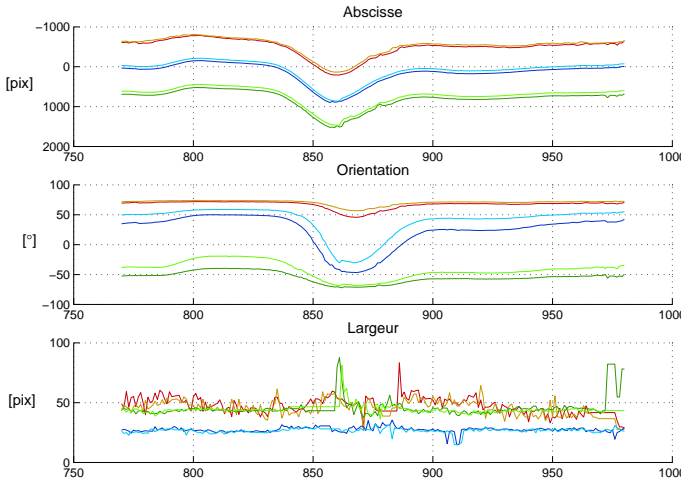


Fig. 4. Chronogram of the road marker characteristics along the Antibes sequence. Although some abscissas of the medians of the three road markers (red left, blue middle, green right) do not belong to the image limits, the filtering process provide reliable predictions which are very close to the estimated location, even in case of large rotation motions (frames [900;1000]).

The last improvements in this stage consist on using a rigid model of a structured road: we obtain relevant results with considering that road markers have constant width along the sequence. Road markers form in image a clear stripe that can

be easily detect if we consider the abscissa and the width at the image bottom along the sequence. A new Kalman filter provides prediction of each stripe abscissa assuming a process with a constant speed. Between the frames [936;952], only one road marker among three is visible when the vehicle has a large rotation motion around the vertice. As soon as the hidden road markers come back in the field of view of the cameras, they are immediately identified (cf Fig. 4).

The Region Of Interest (ROI) that can be assumed as the road plane at the foreground can hence be limited by the two extreme road markers, converging to the DVP, the image bottom and a horizontal. This upper limit represents the region of the road projection assumed to be planed: the ordinate of this line is for the moment chosen as some teens of pixels under the DVP ordinate.

B. Extraction of coplanar features points

1) *The linear system formed by homography:* A homography is a 8 parameters matrix (9-1 of scale factor) which characterizes the motion from a projective plane to a projective plane. According to each feature introduces two equations between its corresponding projections in both plane, only four independent features are required to defined a homography. Whatever the number of coplanar edge-lines we succeed to match, the linear dependency (DVP) between their respective coordinates do not allow a homography computation. We then extract with a Harris corner detector some feature points (FPs) in the ROI.

By definition, the homography computation is strongly linked to the conditioning number of the linear system. That implies the spatial distribution of the matched coplanar features have to be uniform. But due to the scale effect and the road texture, the FPs are mainly extracted in two types of region in the ROI : near the horizon and in the road markers neighborhood. We also face to the paradigm that the FPs at the foreground which best constraint the homography with their longest relative motion between two images are the fewest.

Therefore a novel improvement of the method consist of selecting the five most representative (highest Harris score) FPs in each case of a grid like in Fig 5. This grid split the ROI in two types of sectors: those which contain the road markers and the others. These sectors are next divided in cases with constant height stripes. The grid hence artificially balances the distribution of FPs between regions where many FPs are detected with the others where FPs are extracted from cracks or shadows on the pavement.

Although most of the FPs are extracted in the neighborhood of the road markers, the constraints induced by the road marker medians improve the robustness of the homography computation, according to [12]. Corresponding edge-lines and corresponding FPs do not verify the same system (cf. (1)). The normalization of both sub-systems induced by the matched points and lines is not possible. Best results are nevertheless obtained with only considering the normalization which improves the conditioning number of the sub-system induces by the FPs.

¹<http://www.caor.enscm.fr/~simond>

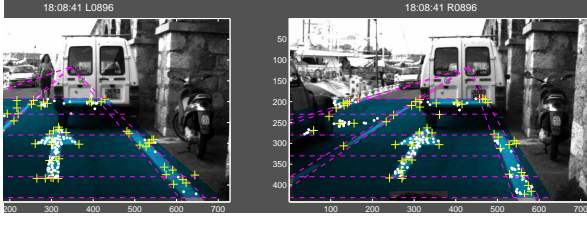


Fig. 5. The grid applied into the blue ROI. Due to the road markers provide the majority of the FPs, the selection of the representative FPs in each case improves the conditioning number of the linear system formed by the FPs projections. Among all the detected FPs '+', only the '+' are selected to compute the current homographies.

2) *The use of stereo-vision:* The use of a stereo-rig makes the identification of the road plane and the extraction of coplanar features easy. According to (3), the stereo homography induced by the road plane depends only on the left camera orientation \mathbf{n}_{st}^k and its altitude d_{st}^k . Furthermore, thanks to the high frame rate of the video sequence, the vehicle motion between two acquisitions is reduced: reliable predictions of the orientation and the altitude of the left camera, assumed as the reference framework, can be provided by their estimations at the last iteration:

$$\hat{\mathbf{n}}_{st}^k \simeq \mathbf{n}_{st}^{k-1} \text{ and } \hat{d}_{st}^k \simeq d_{st}^{k-1} \quad (4)$$

that induces a prediction of the current stereo homography can hence be deduced from the estimated at the last iteration:

$$\hat{H}_{st}^k \simeq H_{st}^{k-1} \quad (5)$$

a) *Matching of edge-lines:* Thanks to the tracking of the road marker medians, we know the filtered coordinates of at least two subsets of coplanar edge-lines in both images. Rather than introduce new couples of unreliable matched edge-lines, homographies are nowadays computed with couples of matched FPs and coordinates of the road markers medians.

b) *Matching of the feature-points:* The knowledge of a prediction of the homography between the two stereo cameras highly simplifies the matching stage which is the crucial task of the whole algorithm. It is based on the fact that only the projections of the road plane satisfy the stereo homography. A first selection consists then to identify couples of coplanar FPs (p_{ls}, p_{rj}) which have reduced re-projection errors with considering the both transformations H_{st} and H_{st}^{-1} :

$$\begin{cases} \text{dist}(H_{st}\mathbf{p}_{ls}, \mathbf{p}_{rj}) < th_dist \\ \text{dist}(\mathbf{p}_{ls}, H_{st}^{-1}\mathbf{p}_{rj}) < th_dist \end{cases} \quad (6)$$

where *dist* represents the euclidean distance and *th_dist* is fixed to 5 pixels.

All the couples of FPs which verify the distance criteria become potential candidates. The ultimate selection consists on computing the normalized correlation $([-1;1])$ between regions $[20 \times 20] \text{ pixels}$ centered on each couple of FPs candidate. Couples whose the covariance correlation is lower than $th_corr = 0.5$ (empirically fixed) can not be assumed as the projections of the same 3D point.

The algorithm of Pilu ([10]), based on the Scott Longuet Higgins method, performs the matching of features in stereo-images with both respect of exclusion and proximity principles. The method relies on the SVD decomposition of an appropriate strength matrix which includes error re-projections and correlation covariance criteria. We personally improve the method robustness with considering a priori knowledge (prediction of H_{st}) on the scene. We also establish a list of corresponding candidates assuming the distance criterion and compute only for them the covariance correlation. The strength matrix is sparse that speeds up the SVD computation and reduces the probability of outliers.

The determination of matched coplanar features requires an iterative process which toggles between the selection of candidate couples with the Pilu method and determination of the induced homography. The number of iterations essentially depends on the spatial distribution of the features. In worst cases, the majority of FPs are aligned along the same road marker: the current method generally failed because the induced linear system is also ill-conditioned. In these case, the homography computation use a random sample consensus RANSAC algorithm ([3]), based on the assumption that the majority of the detected FPs in the ROI belong to the road plane. The number of random selections is determined assuming a fixed rate (5%) of outliers in the sample. This method provides too efficient results in initialization stage but in case of nearly uniform distribution of FPs, the current method is as accurate as the robust one with a shorter time processing.

C. The use of dynamic vision

The homography induced by the road plane between two consecutive images appears less straight forward than the stereo homography because the relative motion of the vehicle is unknown. Most of authors indeed think a feature-based method is not an efficient method to estimate a vehicle motion due to the uncertainty on the number and the reliability of features that can be detected. Nevertheless, assuming that the video frame rate is high, we can then assume that a prediction of the current homography is :

$$\hat{H}_{lr}^k \simeq H_{lr}^{k-1} \quad (7)$$

where H_{lr} represents H_l or H_r .

We also retrieve a similar configuration at the one detailed sooner but this time, we have an a priori knowledge on the coplanar features. The homography computation requires a new iterative process. Nevertheless, there are three cases where the matching stage requires a robust process: at the initialization step, when the prediction of the homography is not reliable or when the spatial distribution of the FPs is not nearly uniform. A coarse prediction of the FPs locations can in fact be estimated assuming that the FPs move down in image with a constant velocity process.

D. Integration of spatio-temporal constraints

The homography computation depends on the coplanar features we achieve to detect. Whatever the type of computed

homography, some couples of FPs located above the road plane are erroneously selected. Our aim is the rejection of these outliers to improve the quality of the homographies. We then verify the coplanarity of each selected FP on different couple of views. That can be resumed as the introduction of spatio-temporal constraints on the feature locations to estimate their ideal coordinates such as the induced homographies verify cross-composition between different views, assuming that the composition rule: $H_{ca} = H_{cb}H_{ba}$ whatever the images \mathcal{I}_a , \mathcal{I}_b and \mathcal{I}_c .

1) *The concept of super-homography*: Malis and Cipolla [8] 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 consistence of the current homographies (H_{st}^k, H_l^k, H_r^k) 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 $m = 6$.

2) *The super feature points*: At the last stages, the computations of current homographies (H_{st}^k, H_l^k, H_r^k) 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 rows 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:

$$\mathbf{p}_{r_s}^k = [H_{st}^k, H_l^k] \cdot \mathbf{p}_{l_s}^{k-1} = [H_r^k, H_{st}^{k-1}] \cdot \mathbf{p}_{l_s}^{k-1} \quad (8)$$

We also introduce a new entity: the super feature point (SFP) which is a $[3m \times 1]$ vector which contains the coordinates of a 3D point in the m last images. If no correspondence between the couples of left or right images ($\mathcal{I}_l^{(k-1)}\mathcal{I}_l^k$ or $\mathcal{I}_r^{(k-1)}\mathcal{I}_r^k$) makes the link between a couple of FPs detected at the last frame, a new SFP is created. Coordinates of unknown projections are completed with zeros.

3) *The virtual feature points*: The pencil of road marker medians is certainly the most reliable feature we succeed to track over the video-sequence. This pencil has the great advantages do not depend on the clear view of the scene and it constraints furthermore the road plane projection in images, mainly at the foreground. We therefore introduce the virtual

feature points (VFPs) to constraint virtually the road plane projection in areas where no SFPs are detected.

The VFPs are then issued from the intersections of the pencil of road marker medians with virtual lines lying on the road, defined by couples of coplanar SFPs whose the coordinates are known 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 $[-45^\circ; 45^\circ]$ with a step of 10° , like in Fig. 6.

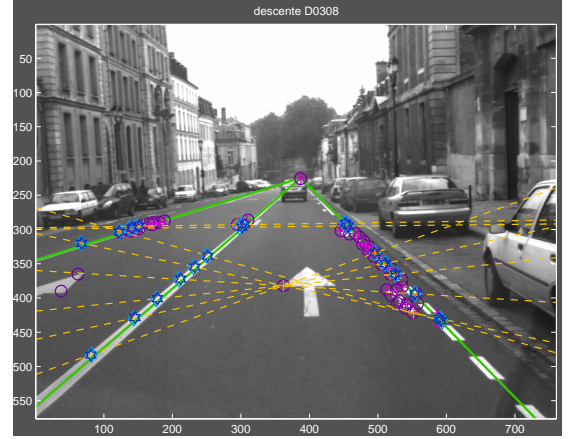


Fig. 6. Virtual feature points (blue ' \diamond ') induced by the intersection of pencil of road markers medians (green lines) and the virtual lines (orange dashed lines), defined with couples of SFPs whose coordinates are known in the m views (orange '+').

4) *The computation of super-homography*: The concatenation of the SFPs and VFPs coordinates in the m images composes a global feature matrix $[3m \times 3n_F]^0 \mathbf{F}$ where $n_F = n_{FP} + n_{VFP}$. Now, let m_r and m_c two integers such as $\{m_r, m_c\} \in [1, m]$ and the induced sequences $\{M_r, M_c\}$ which respectively represent numbers of rows and numbers of columns relative to images \mathcal{I}_{m_r} and \mathcal{I}_{m_c} : $M_\gamma = (3m_\gamma - 2 : 3m_\gamma)$ with $\gamma = \{m_r, m_c\}$.

When the composition rule is verified, the rank of super-homography is:

$$\text{rank}(\mathcal{H}^k) = 3 \quad \forall m \geq 3 \quad (9)$$

Due to all the numerical errors, $\text{rank}({}^1\mathcal{H}^k) > 3$. The first estimation of the n_s^{th} SFP coordinates is also computed according to some projection coordinates are unknown:

$${}^1\mathbf{F}^k(:, n_s) \propto \frac{1}{m^*} {}^1\mathcal{H}^k \cdot {}^0\mathbf{F}^k(:, n_s) \quad (10)$$

where $m^* \leq m$ represents the number of known projections of the n_s^{th} SFP.

Until $q \geq 2$, the process becomes iterative: it toggles with the computation of sub-matrix which form the super-homography from SFPs and the estimation of SFPs coordinates according to the last super-homography computation:

$$\begin{cases} (q-1)\mathbf{F}^k(M_r, :) \propto {}^q\mathcal{H}^k(M_r, M_c) \cdot (q-1)\mathbf{F}^k(M_c, :)^t \\ {}^q\mathbf{F}^k \propto \frac{1}{m} {}^q\mathcal{H}^k \cdot (q-1)\mathbf{F}^k \end{cases} \quad (11)$$

The process stops when $\text{rank}({}^q\mathcal{H}^k) = 3$, 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 locations of SFPs. The ideal sub-pixellic homogeneous coordinates of the SFPs in the current left and right images can be extracted from ${}^q\mathbf{F}^k(3m - 5 : 3m, :)$ whereas the current homographies can be extracted from the final estimation of \mathcal{H}^k :

$$\mathcal{H}^k = \begin{bmatrix} \mathbf{I}_3 & (\mathbf{H}_{\text{st}}^{k-2})^{-1} & \dots & \dots & \dots & \dots \\ \mathbf{H}_{\text{st}}^{k-2} & \mathbf{I}_3 & \dots & \dots & \dots & \dots \\ \dots & \dots & \mathbf{I}_3 & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \mathbf{I}_3 & \vdots & \vdots \\ \dots & \dots & \mathbf{H}_l^k & \dots & \mathbf{I}_3 & (\mathbf{H}_{\text{st}}^k)^{-1} \\ \dots & \dots & \dots & \mathbf{H}_r^k & \mathbf{H}_{\text{st}}^k & \mathbf{I}_3 \end{bmatrix} \quad (12)$$

with $p_s^t = [\mathbf{p}_{\text{ls}}^{k-2}, \mathbf{p}_{\text{rs}}^{k-2}, \dots, \mathbf{p}_{\text{ls}}^k, \mathbf{p}_{\text{rs}}^k]$. Note that the upper matrix of \mathcal{H}^k is the inverse transpose of the lower part.

IV. RECONSTRUCTION OF THE ROAD PLANE

We present in Fig. 7 the reconstruction of the road plane along a video sequence recorded in a Versailles street. The sequence represents a forward motion of the vehicle at abroad 50 km/h along more than 200m with a change of lane at the end. The frame rate was 15 Hz. The bird eye view of the road is 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 also do not constraint sufficiently the super-homography which is finally incorrect. Nevertheless, a temporal filtering process cancel this erroneous estimation.

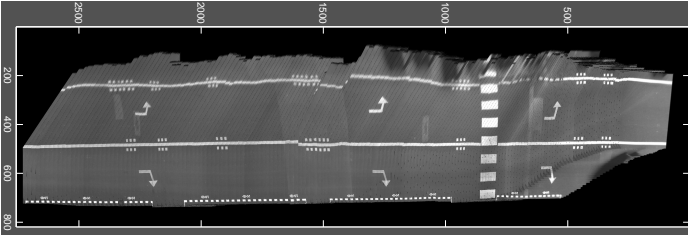


Fig. 7. Reconstruction of the road plane in a Versailles street.

V. CONCLUSION AND PERSPECTIVES

The feature-based method we present allow the reconstruction of the road plane in urban traffic conditions thanks to the data provided by an uncalibrated stereo-rig. The method requires a structured road where the different lanes can be easily identified with the estimation of the DVP. The use of the stereo-rig allows us to reject outliers: couples of matched FPs between two consecutive images which verify the selection criteria due to their low altitudes on the road plane. The methods hence copes with occluded regions by dynamical obstacles. The computation of the super-homography introduces spatio-temporal constraints which improves drastically the matching

process between coplanar features and the conditioning number of the induced homographies.

The last picture shows we succeed to estimate reliability all the homographies induced by the road plane in the right camera along the sequence. That implies the method allows an accurate localization of a vehicle according to the road plane with only the observation of the road at the foreground. This work may have many applications for the urban policy: the pavement service, the cadastre, some real-time navigation systems providers. Furthermore, direct measures are indeed affordable in hybrid image, that offers new solutions for these users.

REFERENCES

- [1] M. 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)*, Head Island, SC, USA, June 13-15 2000, pp. 282–289.
- [2] 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.
- [3] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," in *ACM*, June 1981, pp. 381–395.
- [4] A. Georgiev and P. Allen, "Localization methods for a mobile robot in urban environments," in *IEEE Transactions on Robotics and Automation (ICRA'04)*, 2004.
- [5] 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.
- [6] 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)*, Las Vegas, Nev., USA, Oct. 27-31 2003, pp. 2974–2979.
- [7] Q. Ke and T. 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)*, Madison, USA, June 2003, pp. 390–7.
- [8] 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)*, vol. 2, Dublin, EIRE, June 2000, pp. 610–624.
- [9] 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)*, Washington DC, USA, May 2002, pp. 3321–3327.
- [10] 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)*, San Juan, Puerto Rico, June 17-19 1997, pp. 261–266.
- [11] E. Royer, M. Lhuillier, 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.
- [12] S. Se and M. Brady, "Road feature detection and estimation," *Machine Vision and Applications*, vol. 14, no. 3, pp. 157–165, 2003.
- [13] N. Simond and P. Rives, "Homography from a vanishing point in urban scenes," in *IEEE RSJ/International conference on Intelligent Robot and System (IROS'03)*, Las Vegas, Nev., USA, Oct. 27-31 2003.
- [14] G. Stein, O. Mano, and A. Shashua, "A robust method for computing vehicle ego-motion," in *IEEE Intelligent Vehicle Symposium (IV'00)*, Piscataway, NJ, USA, Oct. 3-5 2000, pp. 362–368.
- [15] 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)*, Las Vegas, Nev., USA, Oct. 27-31 2003, pp. 1308–1313.