HAL
open science

Vehicle Trajectory from an Uncalibrated Stereo-Rig with Super-Homography

Nicolas Simond, Claude Laurgeau

To cite this version:

Nicolas Simond, Claude Laurgeau. Vehicle Trajectory from an Uncalibrated Stereo-Rig with Super-Homography. International Conference on Intelligent Robots and Systems (IROS'06), Oct 2006, Beijing, China. inria-00091212v2

HAL Id: inria-00091212

https://inria.hal.science/inria-00091212v2

Submitted on 23 Oct 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.

Vehicle Trajectory from an Uncalibrated Stereo-Rig with Super-Homography

Nicolas SIMOND, Claude LAURGEAU
Centre de Robotique, Ecole Nationale Supérieure des Mines de Paris,
60 Bd. Saint-Michel, 75 272 Paris Cedex 06, France
Email: Firstname.Lastname@ensmp.fr

***Abstract*— We present in this article an original manner to estimate the trajectory of a vehicle running in urban-like areas. The method consists in extracting then tracking features (points, lines) with an uncalibrated stereo-rig from the road assumed as a plane to compute homographies relative to the camera(s) motions. 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 ([18]) is used in case of unstructured environment or textureless environment,
- direct method ([7], [17]) 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 transposition 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 the concept of super-homography which introduces spatio-temporal rigidity constraints between several views of a planar structure. The third section details how segment the road plane and extract coplanar features. The global methodology is discribed in section four and show in section five the trajectory of a vehicle in urban environment estimated with an uncalibrated stereo-rig.

II. Mathematical background

A. The 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. Each couple of matched features introduces two equations (**u,v** axis) between their respective coordinates in both plane. Hence, only four independent features are required to define a homography. In the following, we generally focus on homographies computed from sets of matched feature points to develop our purpose. The best homography computations are obtained from sets of matched edge-lines and feature points ([14]) although the two types of features do not verify the same projective equalities:

$$\mathbf{p_b} \propto \mathsf{H_{ba}}\mathbf{p_a} \quad \text{and} \quad \mathbf{l_b} \propto \mathsf{H_{ba}^{-t}}\mathbf{l_a} \tag{1}$$

where $\mathbf{p}_\gamma$ and $\mathbf{l}_\gamma$ are respectively the projections in the image $\mathcal{I}_\gamma$ of a 3D point $\mathbf{P}$ and a 3D line $\mathbf{L}$, lying on the π plane.

The recovery of the transformation $[\mathsf{R_{ba}}, \mathbf{t_{ba}}]$ between the two camera(s) poses from the homography $\mathsf{H_{ba}}$, requires the

knowledge of the upper triangular matrix $\mathsf{K_a}, \mathsf{K_b}$ 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 $\mathbf{C_a}$:

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

B. The concept of Super-Homography

1) Introduction: Malis and Cipolla [8] described an efficient method to impose the constraints between the homographies computed from different views of a planar sequence. The concept of *Super-Homography* ($\mathcal{H}$) is indeed based on the strong constraint induced by the projections of a 3D point $\mathbf{P}$, lying on a π plane, in three different views $\mathcal{I}_a, \mathcal{I}_b$ and $\mathcal{I}_c$, like in Fig.1:

$$\mathbf{p_c} \propto \mathsf{H_{cb}}\mathbf{p_b} \propto \mathsf{H_{cb}}\mathsf{H_{ba}}\mathbf{p_a} \propto \mathsf{H_{ca}}\mathbf{p_a} \tag{3}$$

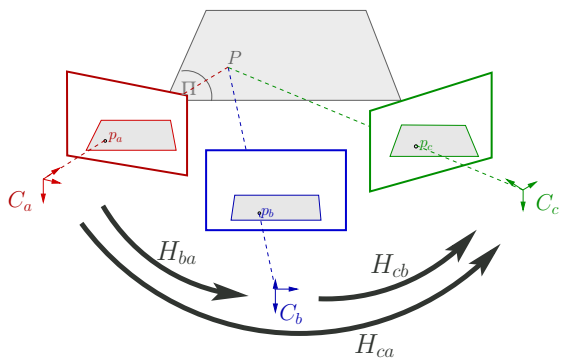

Fig. 1. *The projections of the point* $\mathbf{P}$ *are cross-linked thanks to the homographies induced by the* π *plane in each image.*

These projective equalities are never verified when the homographies are computed with real data. The $\mathcal{H}$ computation consists on determining the theorical coordinates of the matched features in the different views to finally obtain homographies which verify the composition rule (3). When the $\mathcal{H}$ computation ends, features whose coordinates are far from the theoretical ones are not in correspondence or do not lie on the considered plane. The $\mathcal{H}$ computation has also three main advantages:

- it allows the rejection of features which are not lying on the plane,
- it avoid the calibration and the rectification of camera(s) stages because the coordinates of the features which verify $\mathcal{H}$ are also known with a sub-pixellic resolution whatever the image distorsions,
- it reduces numerical instability with the introduction of features coordinates which are temporally occluded or out of field of view in some images.

Thanks to the link between the sets of matched features and the induced homographies, the homographies extracted from $\mathcal{H}$ are most reliable and accurate than those classically computed between each couple of views.

2) The computation of super-homography: Lets now consider m views of a planar structure. The super-homography is a $[3m \times 3m]$ matrix which contains all the $[3 \times 3]$ homographies induced by the plane in the $(m(m-1))$ couples of different views. As soon as the relationships between the projections of 3D features in the different views are identified, the homogeneous coordinates of the m projections of each 3D point are ordered in a column vector that represents a *Super Feature Point* (SFP). The concatenation of all the SFPs form the *Feature* $\mathcal{F}$ matrix ($[3m \times n_{SFP}]$) where coordinates of unknown projections are reset with zeros.

$$\mathcal{H} = \begin{bmatrix} \mathsf{I_3} & \mathsf{H_{12}} & \cdots & \mathsf{H_{1m}} \\ \mathsf{H_{21}} & \mathsf{I_3} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{H_{m1}} & \cdots & \cdots & \mathsf{I_3} \end{bmatrix} \text{ with } \mathcal{F}(:,s) = \left| \begin{matrix} \mathbf{p_{s1}} \\ \mathbf{p_{s2}} \\ \vdots \\ \mathbf{p_{sm}} \end{matrix} \right. \tag{4}$$

The super-homography computation consists on an iterative process which toggles with the estimations of ${}^q\mathcal{H}$ and ${}^q\mathcal{F}$:

$$\begin{cases} {}^{(q-1)}\mathcal{F}(M_r,:) \propto {}^q\mathcal{H}(M_r,M_c).{}^{(q-1)}\mathcal{F}(M_c,:) \\ {}^q\mathcal{F} \propto \frac{1}{m}\,{}^q\mathcal{H}.{}^{(q-1)}\mathcal{F} \end{cases} \tag{5}$$

where $\{M_r, M_c\}$ are indices of $\mathcal{F}$ and $\mathcal{H}$ lines relative to the $\{\mathcal{I}_r, \mathcal{I}_c\}$ views. The process stops as soon as the composition rule between the homographies which form ${}^q\mathcal{H}$ is verified:

$$rank({}^q\mathcal{H}) = 3 \quad \forall m \geq 3 \tag{6}$$

That is generally the case after $q = 3$ iterations unless the system is ill-conditioned due to errors on feature matching or to a non-uniform spatial distribution of the SFPs.

The initialization of both entities is provided from the preceding stages which consists on matching the features sets while computing the homographies between all the couples of views. Only $(m(m-1))/2$ homography computations are required due to the fact that the lower part of $\mathcal{H}$ is the inverse transpose part of the upper one.

The initialization of the feature matrix ${}^0\mathcal{F}$ is computed although some projections coordinates are unknown:

$${}^0\mathcal{F}(:,s) \propto \frac{1}{m^*}\,{}^0\mathcal{H}.{}^0\mathcal{F}(:,s) \tag{7}$$

where $m^* \leq m$ represents the number of known projections of the sth SFP.

III. Application to the Motion of an Uncalibrated Stereo-Rig Motion in Urban Environments

The super-homography computation caters off for the estimation of a stereo-rig motion from a plane where some features can be extracted. It is effectively the case with an embedded stereo-rig which uses the clear view of the road to estimate induced homographies along its path following. In urban environments, the road is generally structured: road makers limit each navigation lane and the road boundaries are highlighted by soft shoulders, green strips or kerbs. The main difficulty of this environment is the dynamic character of the

scenes where the dense traffic strongly reduces the free space from obstacles to the first ten meters before the vehicle.

Hence, urban road can be assumed as a local plane limited by parallel boundaries where RMs, cracks on the pavement and shadow effects allows generate feature points that allows an homography computation between two poses of a camera, modeled with pinhole model. We will now explain how to segment the road plane in the image to extract coplanar features on the road. We will then detail recommendations to compute reliable homographies in these specific conditions to finally justify the use of stereo-vision.

A. Segmentation of the road plane

We detailed in a previous article [15] how to extract the road plane boundaries in images using edges provided by a Canny edge detector. We introduced the *Dominant Vanishing Point* (DVP) which is the intersection between all the 3D lines parallel to the road direction and the retinal plane. The relative motion of the plane is a priori reduced between two poses and most of the majority of edges extracted at the foreground are aligned with the road direction. We also 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 them first to form the projections of the 3D lines in the images.

In the top of Fig. 2, we show the result of the DVP location in both images along the Antibes sequence. The vehicle follows a straight path, stops before a parked vehicle, overtakes it then continues its straight motion. The curves represent the coordinates of the DVP estimated by the Kalman filter in both images. The smoothness of the time diagrams allows us to confirm our assumptions and models (invariant process for ordinate and constant speed evolution for abscissa). Animations on the sequences are available in our web site[1].

The DVP estimation is clearly linked to a stage that corresponds to the tracking of the *Road Markers* (RMs). They indeed form in image a pencil of clear stripes that can be easily detect if we consider the abscissa and the width at the image bottom. According to the RMs have width which slowly evolve along the sequence, characteristics of the RMs measured at the image bottom have relevant properties. We effectually note that each RM median follows a process with a constant speed. A second Kalman filter hence allows us to track the RM characteristics even tough they are out of the field of view. That is the case for the middle and the right RM in both views in the Fig. 2 between the frames [936;952] while the vehicle has a large rotation motion around the vertical. As soon as the hided RMs come back in the field of view of the cameras, they are correctly identified.

The *Region Of Interest* (ROI) that can be assumed as the road plane at the foreground can hence be limited by the two extreme RMs, 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

[1] $http://www-rocq.inria.fr/who/Nicolas.Simond$

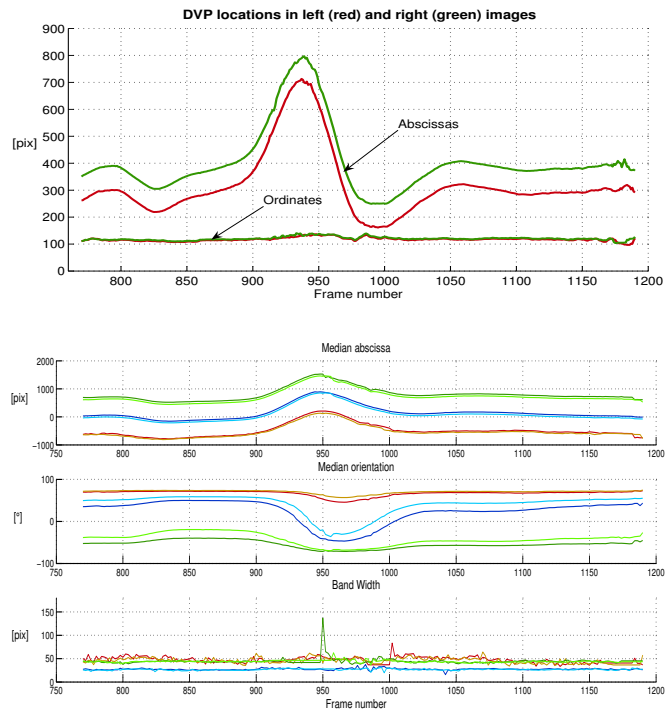

Fig. 2. *Evolution of the DVP and the characteristics of the RMs along the Antibes sequence. Top: chronogram of the DVP location in left (red) and right (green) images. Bottom: tracking of the three RMs (red left, blue middle, green right) RMs for the 2-lanes road.*

for the moment chosen as some teens of pixels under the DVP ordinate.

B. Extraction of coplanar features

At the image foreground (first meters before the test vehicle), the ROI is supposed to be free from obstacle. The *Feature Points* (FPs) we extract in the ROI thus have a high probability to lie on the road. The FPs detected on the road area have two drawbacks: they have poor characterization and have generally non-uniform spatial distribution. Therefore most of authors do not expect obtain relevant results with such a method for this type of application.

However, the homography computation requires only the matching of at least four features in both images. According to the tracking of RMs already provides reliable couples of matched features, the matching of the couples of FPs which represent projections of 3D points is based on a variant of the *Iterative Closest Point (ICP)* alignment algorithm [12]. The method toggles the computation of homography with the matching of features between the both images.

The algorithm of Pilu ([10]), based on the Scott Longuet-Higgins ([13]) method, is dedicated to the matching of stereo images (small baseline). An appropriate correspondence strength matrix is performed at each iteration: it integrates both a geometric (inter-distance between the two clouds of FPs) and an intensity relationship (correlation between gray level values of patchs centered on each FP). The SVD decomposition of this matrix provides couples of FPs which both respects the

exclusion and proximity principles. We implement the method in [16].

Some couples of outliers generally appear when the induced homography is not sufficiently constrained (distribution of features non-uniform). A RANSAC method ([3]) is also performed at initialization stage or as soon as the spatial distribution of features is non-uniform.

C. Homography induced by the road plane

By definition, a homography is a linear system which makes the link between homogeneous coordinates of two sets of matched features. In our case, we use camera(s) fixed behind the windshield of a car to detect features (points and lines) lying on the road. Road generally appears as textureless surface where most of the detected FPs are extracted around or on the RMs. Moreover, due to the perspective effect, the FPs which are best characterized (far from the camera) have the worst resolution whereas only the features detected on the foreground best constraint the camera motion because they have the longest relative motion. At least, the matching stage of FPs with low score is unreliable.

The lonely way to compute accurate homographies is to perform their conditioning number. In other words, the features matched between two poses have to be uniformly distributed in the ROI. We also use a grid applied to the ROI (cf. Fig. 3) to impose an almost-uniform distribution for the FPs. Only the most representative (highest Harris score) FPs of them are preselected. The grid hence artificially balances the distribution of FPs in the ROI in a similar area that has the great advantage do not introduce a crude threshold to select the FPs and reject most of the FPs detected in the RMs which have low characterization.

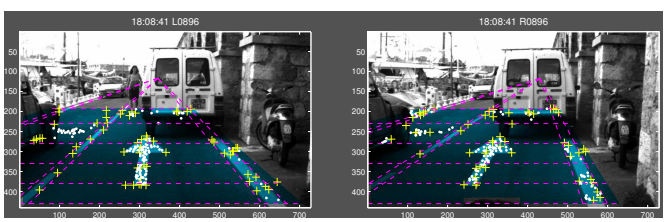

Fig. 3. *The grid applied into the blue ROI is related to the 2-lane road model. Among all the FPs '.' detected, only the most representative of each case ('+') are selected to compute the current homographies.*

IV. Global methodology

A. Outlines

The estimation of the stereo-rig motion between two poses can be divided into four linked sequential stages according to the synoptic of Fig.4 left. 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 are extracted according to the stereo constraints. These couples have been tracked between two consecutive images during the third stage. The final stage provides robust coplanar features with verifying the coplanarity of features detected in the current couple of images with the last one(s).

The stereo-rig motion at the current frame (k) mainly infers three camera motions (cf Fig. 4 right): the stereo homography (blue $\mathsf{H}_{\mathsf{st}}^{\mathsf{k}}$) to verify the rigid transformation between both cameras, the left (red $\mathsf{H}_{\mathsf{l}}^{\mathsf{k}}$) and right (green $\mathsf{H}_{\mathsf{r}}^{\mathsf{k}}$) camera motions. These homographies are computed according to their reverse motion (dashed arrows). The $\mathcal{H}^k$ computation introduces strong rigidity constraints with considering all the preceding homographies plus the cross-linked ones (purple) between the camera poses.

B. Use of stereo-vision

Thanks to the stereo-rig is rigidly fixed behind the windshield, we assume that the transformation between the left and right camera frameworks is rigid:

$$\mathsf{R}_{\mathsf{st}}^{\mathsf{k}} \equiv \mathsf{R}_{\mathsf{st}}^{\mathsf{k-1}} \quad \text{and} \quad \mathbf{t}_{\mathbf{st}}^{\mathbf{k}} \equiv \mathbf{t}_{\mathbf{st}}^{\mathbf{k-1}} \tag{8}$$

The *stereo homography* H_{st} hence depends only on the left camera orientation $\mathbf{n}_{\mathbf{st}}^{\mathbf{k}}$ and its height 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 height of the left camera, assumed as the reference framework, can be provided by their estimations at the last iteration:

$$\widehat{\mathbf{n}}_{\mathbf{st}}^{\mathbf{k}} \simeq \mathbf{n}_{\mathbf{st}}^{\mathbf{k-1}} \quad \text{and} \quad \widehat{d}_{st}^k \simeq d_{st}^{k-1} \tag{9}$$

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

$$\widehat{\mathsf{H}}_{\mathsf{st}}^{\mathsf{k}} \simeq \mathsf{H}_{\mathsf{st}}^{\mathsf{k-1}} \tag{10}$$

The use of a stereo-rig hence makes easier the identification of the road plane and the extraction of coplanar features.

C. Use of dynamic vision

According to the slow motion assumption between two consecutive frames, a similar approach leads to a coarse prediction for homographies can be: $\widehat{\mathsf{H}}_{\mathsf{lr}}^{\mathsf{k}} \simeq \mathsf{H}_{\mathsf{lr}}^{\mathsf{k-1}}$ where $_{lr}$ represents either the left or right camera. However, the motion between two consecutive poses is unknown but the continuity of the vehicle forward motion implies a radial motion of the road plane. This last remark allows us to constrain the matching stage of the FPs because the magnitude of the FP relative motion depends on their distance to the camera.

Thanks to the coarse prediction of homography, we also retrieve a similar configuration at the stereo one but this time, we have an a priori knowledge on the coplanar features. The H_{lr}^k homograhy computation also requires a similar iterative process as the one detailed for stereo-homography.

D. Spatio-temporal filtering

The first three stages of the method can be viewed as a global initialization of the fourth. The super-homography actually requires sets of matched features between all couples of the m views. In case of using a stereo-rig, we assume that the SFPs used to compute $\mathcal{H}^{k-1}$ are correct. We also only

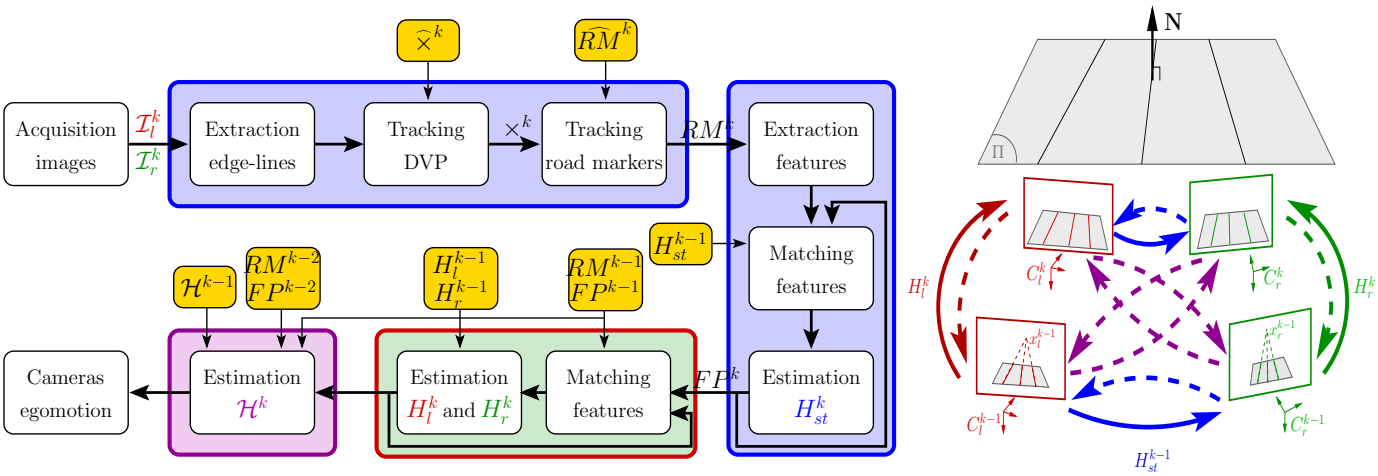

Fig. 4. *Synoptic of the whole method. On the left, the different stages required to compute $\mathcal{H}^k$ from the images $\mathcal{I}_l^k$ and $\mathcal{I}_r^k$. The first two stages correspond to the identification of coplanar features (DVPs, RMs, FPs) in the current images while the third make the link with the last frames. On the right, the relative motion of the stereo-rig between the last $(k-1)$ and current (k) poses mainly implies the homographies $\mathsf{H}_{\mathsf{st}}^{\mathsf{k}}, \mathsf{H}_{\mathsf{l}}^{\mathsf{k}}$ and $\mathsf{H}_{\mathsf{r}}^{\mathsf{k}}$.*

have to determine what are the new projections of these SFPs in the current couple of images ($\mathcal{I}_l^k$ and $\mathcal{I}_r^k$). We find them with the computation of the homographies $\mathsf{H}_{\mathsf{st}}^{\mathsf{k}}, \mathsf{H}_{\mathsf{l}}^{\mathsf{k}}, \mathsf{H}_{\mathsf{r}}^{\mathsf{k}}$.

Due to the spatial distribution of the SFPs is generally not uniform. We also introduce a second type of SFPs which drastically increase the uniformity of the spatial distribution in the ROI: the Virtual Feature Points.

1) The Virtual Feature Points: The pencil of RM is certainly the most reliable feature we succeed to track over the video-sequence. Its detection is robust with regard to the local occlusions due to the presence of obstacles on the road. Thanks to this property, we can define a new type of features, called *virtual feature points (VFPs)*, to constrain the road plane projection in areas where no SFP is detected.

The VFPs are built from the intersections of the pencil of RM medians with virtual lines, defined by couples of coplanar SFPs observed in all 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 distributed between $[-45°; 45°]$ with a step of $10°$.

We present in Fig. 5 the warping of the current image in the last one with considering the homography computed classically (on top) and extracted from $\mathcal{H}^k$ (on bottom). The spatio-temporal constraints introduced by the VFPs and the induced $\mathcal{H}^k$ also improve the re-projection of the road plane.

2) Selection of the m parameter: The determination of the optimal number of views (m) to compute $\mathcal{H}$ is not trivial. To cast off ambiguities between rotation and translation motions during the homography decomposition stage, the relative motion of the plane in images have to be significant. A raise of the distance between views has still the drawback to reduce the rigidity constraints: less features are common to all views. We personally use three stereo couples corresponding to the frames $(k-2)$, $(k-1)$ and (k) but an adaptive selection stage should be implemented to optimize m with the vehicle speed.

V. Vehicle Trajectory

The intrinsic parameters of both cameras are computed with an off-line calibration based on a grid pattern. The rotation and translation motion of the camera which induces the homography can be retrieved from (2). Due to the fact that homography depends on 8 parameters, its decomposition requires an a priori knowledge of the normal to plane to distinguish the right solution among the two transformation couples which rely on **n** and (**-n**). In our configuration, the (**z**) optical axis is nearly parallel to the vehicle axis. The normal to plane is also near to the vertical axis ($\mathbf{y} = [0; -1; 0]^t$). The discontinuities of the translation along the **z** axis on top of Fig. 6 are only due to a sudden increase of the time-stamp.

The Versailles sequence represents a forward motion of the vehicle at abroad 50 km/h along more than $350m$ with a change of lane before stopping. The frame rate is approximately 15 Hz. The bird eye view of the road (Fig. 6 bottom) is made of the warping of 310 images of the ROI segmented in the right images of the stereo-rig. Due to the common parallel axis of the cameras, when the vehicle changes of lane, the common region between the left and right images decreases. The extracted FPs are also not sufficiently distributed in the ROIs to correctly constraint the homographies computation.

VI. Conclusion

The spatio-temporal constraints introduced by the projections of 3D features in different views allows the rejection of outliers and improves the quality of induced homographies with respecting the composition rule. The introduction of Virtual Feature Points authorizes the computation of super-homographies where the spatial distribution of the matched features is at least almost-uniform. This method provides high quality results even in case of occlusions on the road plane. We achieved to reconstruct the road over hundreds of meters with coplanar features detected with an uncalibrated stereo-

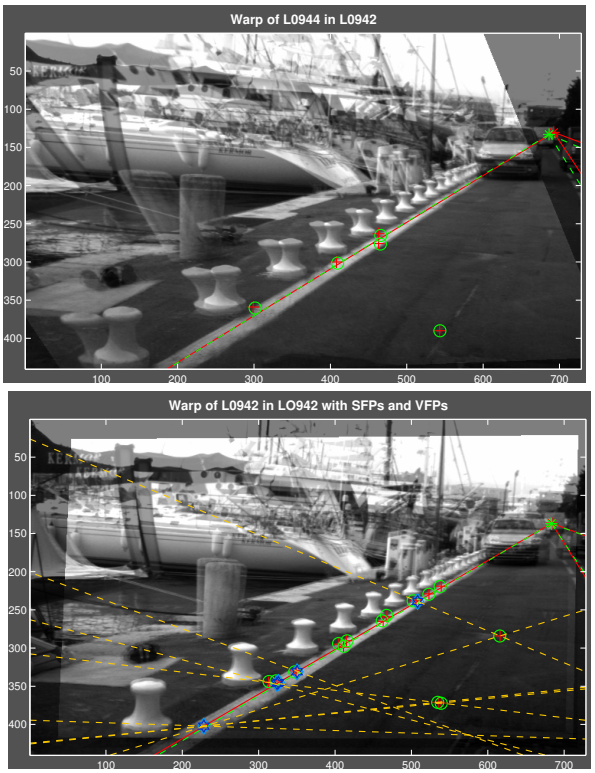

Fig. 5. *Comparison between homography computed classically (top) and that extracted from the super-homography (bottom). The current image with green features is warped into the last one with red features. The VFPs (blue '◇') are intersections of pencil of RMs and the virtual (orange dashed) lines.*

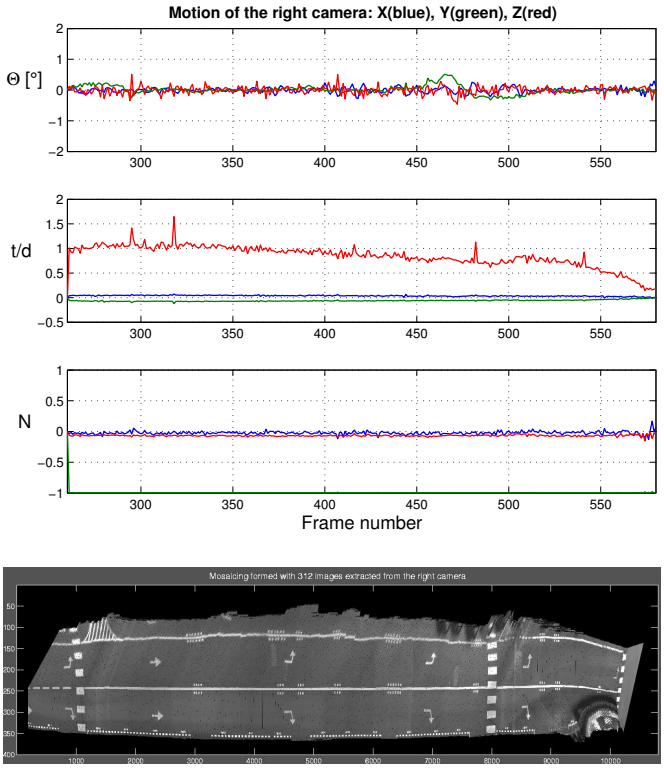

Fig. 6. *Versailles sequence. Top: Decomposition of the right camera homographies induced by the road plane: from top to bottom we retrieve rotation then translation motions and the orientation of the normale plane. Bottom: bird eye view of the road with right images between frames [260;570].*

rig. The reconstruction distance is representative of the length where no GPS data can be received in a urban canyon.

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. 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.

[12] S. Rusinkiewicz, "Rigid-body alignment," in *IEEE International Conference on Computer Vision (ICCV'05)*.

[13] G. Scott and H. Longuet-Higgins, "An algorithm for associating the features of two patterns," in *Royal Society of London*, vol. B244, 1991, pp. 21–26.

[14] S. Se and M. Brady, "Road feature detection and estimation," *Machine Vision and Applications*, vol. 14, no. 3, pp. 157–165, 2003.

[15] 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.

[16] ——, "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.

[17] 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.

[18] 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.