Real-Time Monocular Visual Odometry for On-Road Vehicles with 1-Point RANSAC

Davide Scaramuzza[1], Friedrich Fraundorfer[2], and Roland Siegwart[1]

[1]Autonomous Systems Lab, ETH Zurich
[2]Computer Vision and Geometry Group, ETH Zurich
davide.scaramuzza@ieee.org, fraundorfer@inf.ethz.ch, r.siegwart@ieee.org

Abstract— **This paper presents a system capable of recovering the trajectory of a vehicle from the video input of a single camera at a very high frame-rate. The overall frame-rate is limited only by the feature extraction process, as the outlier removal and the motion estimation steps take less than 1 millisecond with a normal laptop computer. The algorithm relies on a novel way of removing the outliers of the feature matching process. We show that by exploiting the nonholonomic constraints of wheeled vehicles it is possible to use a restrictive motion model which allows us to parameterize the motion with only 1 feature correspondence. Using a single feature correspondence for motion estimation is the lowest model parameterization possible and results in the most efficient algorithms for removing outliers. Here we present two methods for outlier removal. One based on RANSAC and the other one based on histogram voting. We demonstrate the approach using an omnidirectional camera placed on a vehicle during a peak time tour in the city of Zurich. We show that the proposed algorithm is able to cope with the large amount of clutter of the city (other moving cars, buses, trams, pedestrians, sudden stops of the vehicle, etc.). Using the proposed approach, we cover one of the longest trajectories ever reported in real-time from a single omnidirectional camera and in cluttered urban scenes, up to 3 kilometers.**

I. INTRODUCTION

While there exist nowadays a wide availability of algorithms for motion estimation using video input alone (see Section II), cameras are still little integrated in the motion estimation system of a mobile robot and even less in that of an automotive vehicle. The main reasons for this are the following:

- several algorithms can still only work off-line or at low frame-rate,
- others need high processing power or expensive and dedicated processors,
- many algorithms are quite complex to use or are designed for specific cameras,
- many algorithms assume static scenes and cannot cope with dynamic and cluttered environments or huge occlusions by other passing vehicles (like what happens in typical urban environments in real traffic with other moving cars, buses, trams and pedestrians, sudden changes of speed, etc.),
- the data-association problem (feature matching and outlier removal) is not completely robust and can fail,
- the motion estimation scheme usually requires many keypoints and can fail when only a few keypoints are available in almost absence of structure.

Here, we will show that all these areas can be improved by using a restrictive motion model which allows us to parameterize the motion with only 1 feature correspondence. Using a single feature correspondence for motion estimation is the lowest model parameterization possible and results in the most efficient algorithms for removing outliers.

Our approach exploits the nonholonomic constraints of wheeled vehicles, that is, they possess an Instantaneous Center of Rotation (ICR). Cars are typical examples of such vehicles. As everybody experiences in driving, one needs to act on the steering to change the direction of the car. What actually happens in practice is that the two front wheels are turned of a slight different angle to make the vehicle move instantaneously along a circle and, thus, turn about the ICR. As the reader can perceive, this constraint reduces the degrees of freedom of the motion to two, namely the rotation angle and the radius of curvature. The first consequence is that only one feature correspondence suffices for computing the epipolar geometry. This allows motion to be computed also from scenes where structure is almost absent, provided that at least one feature is available. The second consequence is that very efficient methods for removing outliers can be implemented.

Note, in this paper we will show experiments with an omnidirectional camera but the proposed approach can be applied also to perspective cameras.

The outline of the paper is the following. In Section II, we review the related work. In Section III, we explain how the nonholomic constraints of wheeled vehicles can be used to compute the calibrated essential matrix from one point. In Section IV, we describe two methods for outlier removal. In Section V, we summarize our motion estimation algorithm. Finally, in sections VI and VII we present our experimental results and conclusions.

Please observe that the following paper is accompanied by a demonstrative video available from the author's webpage.

II. RELATED WORK

The problem of recovering relative camera poses and 3D structure from a set of monocular images has been largely studied for many years and is known in the computer vision community as "Structure From Motion" (SFM) [1]. Successful results with only a single camera and over long distances (from hundreds of meters up to kilomeers) have been obtained in the last decade using both perspective and omnidirectional cameras (see [2]–[9]). Here, we review some of these works.

Related works can be divided into three categories: feature based methods, appearance based methods, and hybrid methods. Feature based methods are based on salient and repetitive features that are tracked over the frames; appearance based methods use only the intensity information of the whole image or subregions of it; hybrid methods use a combination of the previous two.

In the first category are the works of [2]–[5]. In [2], Bosse *et al.* used vanishing points and 3D lines to recover both structure and motion on a 946-meter path. In [5], Nister *et al.* dealt with the case of a stereo camera but they also provided a monocular solution implementing a full structure from motion algorithm that takes advantage of the 5-point algorithm and RANSAC robust estimation. In [3], Corke *et al.* provided two approaches for monocular visual odometry based on omnidirectional imagery from a catadioptric camera. As their approach was conceived for a planetary rover, they performed experiments in the desert and therefore used keypoints from the ground plane. In the first approach, they used optical flow computation with planar motion assumption while in the second one they performed full SFM (with unconstrained motion). The optical flow method with planar assumption gave the best performance over 250 meters but the trajectory was not accurately recovered showing a large drift of the rotation estimation. In [4], Lhuillier used 5-point RANSAC and bundle adjustment to recover both the motion and the 3D map. In [7], Tardif *et al.* presented an approach for incremental and accurate SFM from a car over a very long run (2.5 km) without bundle adjustment. To achieve it, they decoupled the rotation and translation estimation. In particular, they estimated the rotation using points at infinity and the translation from the recovered 3D map. Bad correspondences were removed with preemptive RANSAC [10]. However, in contrast with the previous works, they did not use a catadioptric camera but a *Ladybug*2 (from PointGrey) which uses six high resolution cameras arranged to give an omnidirectional view.

Among the appearance based or hybrid approaches are the works of [6], [8], [9]. In [6], Goecke *et al.* used the Fourier-Mellin Transform for registering perspective images of the ground plane taken from a car. Results were shown on a 300-meter path. In [8], Milford *et al.* presented a method to extract approximate rotational and translational velocity information from a single perspective camera mounted on
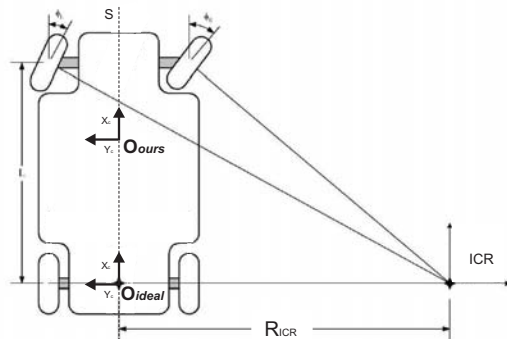

Fig. 1. General Ackermann steering principle (courtesy of Bjorn Jensen).

a car, which was then used in a RatSLAM scheme [11] to generate a coherent map of the urban environment. However, appearance based approaches alone are not very robust to occlusions. For this reason, in our previous works [9], [12] we used appearance to estimate the rotation of the car and features from the ground plane to estimate the translation and the scale factor. The feature based approach was also used as a firewall to detect failure of the appearance based method.

Closely related to structure from motion is what is known in the robotics community as Simultaneous Localization and Mapping (SLAM), which aims at estimating the motion of the robot while simultaneously building and updating the environment map. SLAM has been most often performed with other sensors than regular cameras, however in the last years successful results have been obtained using single cameras alone (see [13], [14], [15], and [16]).

III. WHY DO WE NEED ONLY ONE POINT?

For a wheeled vehicle to exhibit rolling motion, a point must exist around which each wheel of the vehicle follows a circular course [17]. This point is known as Instantaneous Center of Rotation (ICR) and can be computed by intersecting all the roll axes of the wheels (Fig. 1). For cars the existence of the ICR is ensured by the Ackermann steering principle [17]. This principle ensures a smooth movement of the vehicle by applying different steering angles to the inner and outer front wheel while turning. This is needed as all the four wheels move in a circle on four different radii around the ICR (see Fig. 1).

As the reader can perceive, the motion of a camera installed on the vehicle can be then locally described with circular motion; straight motion can be represented along a circle of infinite radius. In the remainder, we will refer to our motion model as "circular motion". Now, we will see how this reflects on the rotation and translation and on the parameterization of the essential matrix. In the following we will assume locally planar motion.

A. The Essential Matrix

Under planar motion, the two relative poses of a camera can be described by three parameters, namely the yaw angle

θ and the polar coordinates (ρ, ϕ) of the second position relative to the first position (Fig. 2). Since when using only one camera the scale factor is unknown, we can arbitrarily set ρ at 1. From this it follows that only two parameters need to be estimated and so only two image points are required. However, if the camera moves locally along a circumference (as in Fig. 2) and the x-axis of the camera is set perpendicular to the radius R_{ICR}, then we have $\phi = \theta/2$; thus, only θ needs to be estimated and so only one image point is required. Observe that straight motion is also described through our circular motion model; in fact in this case we would have $\theta = 0$ and thus $\phi = 0$.

Let us now derive the expression for the essential matrix using the considerations above. Let $\mathbf{R}$ and $\mathbf{T}$ be the unknown rotation and translation matrices which relate the two camera poses. Then, we have

$$\mathbf{R} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ \mathbf{T} = \rho \cdot \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \\ 0 \end{bmatrix} \quad (1)$$

because we considered the motion along the xy plane and the rotation about the z-axis. Then, let $\mathbf{p} = [x, y, z]^T$ and $\mathbf{p}' = [x', y', z']^T$ be the image coordinates of a scene point seen from the two camera positions. Observe that to make our approach independent of the camera model we use spherical image coordinates; therefore $\mathbf{p}$ and $\mathbf{p}'$ are the image points back projected onto a unit sphere (i.e. $\|\mathbf{p}\| = \|\mathbf{p}'\| = 1$). This is always possible once the camera is calibrated.

As known in computer vision, the two unknown camera positions and the image coordinates must verify the epipolar constraint

$$\mathbf{p}'^T \mathbf{E} \mathbf{p} = 0, \quad (2)$$

where E (called *essential matrix*) is defined as $\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$, where $[\mathbf{T}]_\times$ denotes the skew symmetric matrix

$$[\mathbf{T}]_\times = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}. \quad (3)$$

Then, using (1), (3), and the constraint $\phi = \theta/2$, we obtain the expression of the essential matrix for planar circular motion:

$$\mathbf{E} = \rho \cdot \begin{bmatrix} 0 & 0 & \sin(\frac{\theta}{2}) \\ 0 & 0 & -\cos(\frac{\theta}{2}) \\ \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) & 0 \end{bmatrix} \quad (4)$$

B. Recovering θ

By replacing (4) into (2), we can observe that every image point contributes to the following homogeneous equation:

$$\sin\left(\frac{\theta}{2}\right) \cdot (x'z + z'x) + \cos\left(\frac{\theta}{2}\right) \cdot (y'z - z'y) = 0 \quad (5)$$

Given one image point the rotation angle θ can then be obtained from (5) as:

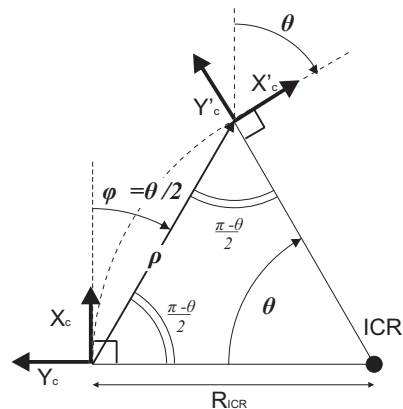$$\theta = -2\tan^{-1}\left(\frac{y'z - z'y}{x'z + z'x}\right) \quad (6)$$

Fig. 2. Relation between camera axes in circular motion.

Conversely, given m image points, θ can be computed indirectly by solving linearly for the vector $[\sin(\frac{\theta}{2}), \cos(\frac{\theta}{2})]$ using Singular Value Decomposition (SVD). To this end, we first form a $m \times 2$ data matrix D, where each row is formed by the two coefficients of Equation (5), that is:

$$\left[(x'z + z'x), \ (y'z - z'y) \right]. \quad (7)$$

The matrix D is then decomposed using SVD:

$$D_{m \times 2} = U_{m \times 2} \Lambda_{2 \times 2} V_{2 \times 2} \quad (8)$$

where the columns of $V_{2 \times 2}$ contain the eigenvectors e_i of $D^T D$. The eigenvector $e^* = [\sin(\frac{\theta}{2}), \cos(\frac{\theta}{2})]$ corresponding to the minimum eigenvalue minimizes the sum of squares of the residuals, subject to $\|e^*\| = 1$. Finally, θ can be computed from e^*.

C. Discussion on our Motion Model

Note, the equations above are valid only when the camera is placed along the back-wheel axis and with the x-axis perpendicular to it (see O_{ideal} in Fig. 1). However in practice cars have small steering angle (Fig. 3), and thus a big radius of curvature. This allows us to relax the previous assumption and to place the camera anywhere above the car provided that the x-axis of the camera is perpendicular to the back-wheel axis. For our car, for instance, the camera was placed as shown in Fig. 1 (O_{ours}) and Fig. 6. This position was chosen arbitrarily without any particular reason but having a wide field of view on the front of the car.

Finally, observe that the planar assumption and the circular motion constraint hold only locally, but because of the smooth motion of cars we found that this assumption actually holds very well also at low frame rates (< 10 Hz while running at 50 Km/h).

IV. OUTLIER REMOVAL: TWO APPROACHES

Here, we describe two approaches for removing the outliers of the feature matching process by using the motion model of the previous section. Once the outliers are identified, the motion estimate can be refined using all the remaining inliers (see Section V). The two approaches explained here are based on RANSAC and histogram voting.
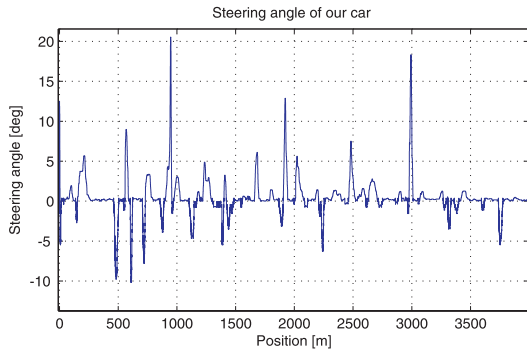
Fig. 3.   Steering angle (*deg*) vs. traveled distance (*m*) read from our car.

TABLE I

Min. set of points:	6 points	5 points	2 points	1 point
No. of iterations:	292	145	16	7

Fig. 4.   Left: side view. Right: top view.

Fig. 5.   An example histogram from feature correspondences.

A. 1-Point RANSAC

The *random sample consensus* (RANSAC) [18] has been established as the standard method for model estimation in the presence of outliers. Structure from motion is one application of the RANSAC scheme. RANSAC works by generating model hypothesis from randomly sampled minimal data sets and verifying them on the whole data set. The number of hypothesis (iterations) N that is necessary to guarantee that a correct solution is found can be computed by $N = \frac{\log(1-p)}{\log(1-(1-\varepsilon)^s)}$, where s is the number of minimal data points, ε is the percentage of outliers in the data points and p is the requested probability of success [18]. N is exponential in the number of data points necessary for estimating the model, so there is a high interest in finding the minimal parameterization of the model. For unconstrained motion of a calibrated camera this would be 5 point correspondences [19], [20]. Using the 6-point algorithm [21] would increase the number of necessary iterations and therefore slow down the motion estimation algorithm. It is therefore of utmost importance to find the minimal parameterization of the model to estimate. In the case of planar motion, the motion model complexity is reduced and can be parameterized with 2 points as described in [22]. For automotive applications we showed in Section III that an even more restrictive motion model can be chosen which allows us to parameterize the motion with only 1 feature correspondence. Using a single feature correspondence for motion estimation is the lowest model parameterization possible and results in the most efficient RANSAC algorithm. Table I shows the number of RANSAC iterations needed for different motion estimation algorithms which require a different number of minimal data points s. These values were obtained assuming a probability of success $p = 99\%$ and a percentage of outliers $\varepsilon = 50\%$. The table shows that with the 1-point RANSAC the needed iterations are extremely low.

In the hypothesis generation step, our 1-point RANSAC computes the relative motion from a single randomly chosen point correspondence. To do this, we use Eq. (6). In the
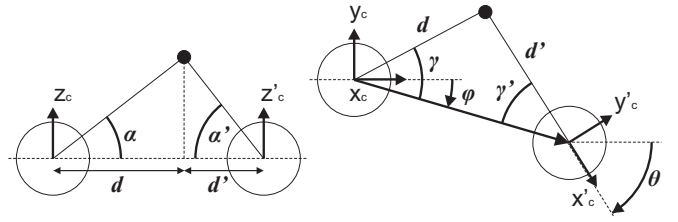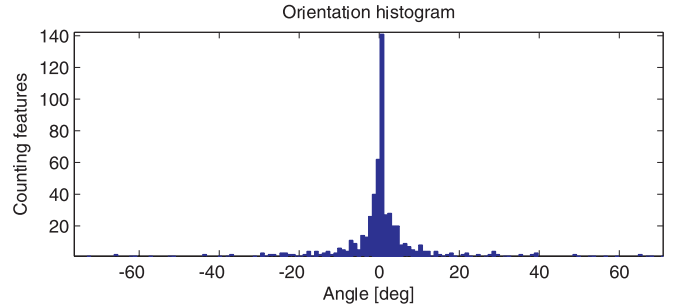
model verification step the consensus set for the model hypothesis is computed, i.e. set of inliers. A directional error measure is used to find the inliers that support the motion model. The error measure is illustrated in Fig. 4. We can observe that for each correspondence we have $d \tan \alpha = d' \tan \alpha'$ and $d \sin \gamma = d' \sin \gamma'$. This implies that for each correspondence the ratio $\frac{d'}{d}$ can be computed both as $\frac{\tan \alpha}{\tan \alpha'}$ and $\frac{\sin \gamma}{\sin \gamma'}$. We consider a corresponding point pair as inlier if the difference of both ratios is lower than a threshold t.

B. Histogram voting

The possibility of estimating the motion using only one feature correspondence allows us to to implement another algorithm for outlier removal that is much more efficient than the 1-point RANSAC approach as it requires no iterations. The algorithm is based on histogram voting: first, θ is computed from each feature correspondence using (6); then, a histogram H is built where each bin contains the number of features which count for the same θ. A sample histogram built from real data is shown in Fig. 5. When the circular motion model is well satisfied, the histogram has a very narrow peak centered on the best motion estimate θ^*, that is $\theta^* = argmax\{H\}$. As the reader can perceive, θ^* represents our motion hypothesis; knowing it, the inliers can be identified as we did the previous section by using again the directional error.

Note, this method implies to compute first the histogram of the θ values in order to then determine θ^*. As a matter of fact, we found an even more efficient solution where, instead of building the histogram, we set θ^* equal to the median of the distribution, that is, $\theta^* = median\{\theta_i\}$. We found the latter giving as good results as the histogram voting.

V. STRUCTURE FROM MOTION ALGORITHM

In the previous sections we introduced our motion model (locally planar circular motion) and we showed that it can be used to remove the outliers of the feature matching process. As we saw, the relative motion between two frames is also computed while removing the outliers. This is done in order to identify the number of inliers supporting the estimated motion. However, our circular motion model (2DoF) is only an approximation of the real motion of the camera. Therefore, we decided to compare it with the planar motion model (3DoF) and the unconstrained motion model (6DoF). The algorithms to estimate structure and motion from a set of inliers with or without planar assumption can be found in [1], [19], [20], [22]. These algorithms proceed similarly to what we did in Section III-B by computing the essential matrix and then decomposing it linearly into $\mathbf{R}$ and $\mathbf{T}$. Observe that the scale (ρ) cannot be recovered from two images of a single camera, therefore in our experiments we used the knowledge of the speed of the vehicle. Our motion estimation algorithm operates as follows:

1) Load a new frame
2) Extract feature correspondences between current and previous frame
3) Compute the pixel distance between the matched points. If more than 90% of the distances is less than 3 pixels then assume no motion and return to step 1
4) Remove the outliers using one of the two methods given in Section IV
5) Recompute motion (θ_c, ϕ_c) from all the remaining inliers using SVD (Section III-B); remember to set $\phi_c = \theta_c/2$ because of the circular motion constraint. Set temporarily the scale factor $\rho = 1$
6) Recompute motion from inliers: use (a) planar assumption or (b) unconstrained motion. Set temporarily the scale factor $\rho = 1$
7) Firewall condition: if the recomputed motion parameters differ too much ($> 10°$) from the θ_c, ϕ_c values of step 5, then reject the recomputed parameters and return θ_c, ϕ_c
8) Get the current speed of the car v and set the image scale $\rho = v\Delta t$
9) Recover the 3D structure by triangulating the rays of the back projected image points of the last two frames
10) Integrate the motion and repeat from step 1

The firewall condition allows us to face those situations with too few features or where the features are distributed on small parts of the image. In these cases we found that the structure from motion algorithms for planar or unconstrained motion returned a motion vector inconsistent with the real movement of the car. Conversely, the circular motion constraint returned always consistent estimates. This is an advantage of exploiting the nonholonomic constraints of wheeled vehicles.

VI. RESULTS

The approach proposed in this paper has been successfully tested on a real vehicle equipped with an omnidirectional

Fig. 6. The vehicle used in our experiments equipped with the omnidirectional camera (in the circle). The vertical field of view is indicated by the lines.

camera. A picture of our vehicle (a Smart) is shown in Fig. 6. Our omnidirectional camera is composed of a hyperbolic mirror (KAIDAN 360 One VR) and a digital color camera (SONY XCD-SX910, image size 640×480 pixels). The camera was installed as shown in Fig. 6. The maximum frame-rate of this camera is 15Hz but in practice we observed always 10Hz. Sometimes the frame-rate dropped to 5Hz because of the memory resources shared with other sensors on the car. For calibrating the camera we used the toolbox described in [23] and available from [24]. The vehicle speed ranged between 0 and 45Km/h.

The dataset was taken in real traffic during the peak time in the city center of Zurich. Therefore, many pedestrians and passing trams, buses, and cars were also present. The images were collected from the beginning until the end of the tour, also when the vehicle was still in the presence of stop signs, pedestrian crossings, and red lights. In the absence of motion, the frames were skipped as explained in Section V. The overall length of the tour is about 3Km and is shown in Fig. 8 overlaid on a satellite image.

We tested our structure from motion algorithm on different feature detectors: SIFT, Harris, and KLT. SIFT returned about $700 \sim 1000$ features per frame, while Harris and KLT about $2500 \sim 4000$ features. We applied these detectors on the same image dataset to enable comparison.

For each feature type we ran the algorithm of Section V. We used independently our two approaches for outlier removal (i.e. 1-Point RANSAC and histogram voting). Both of them perform equally well and can be interchanged.

Then we recomputed the motion from all the inliers using: a) planar circular motion model (2DoF), b) planar motion model (3DoF), c) unconstrained motion model (6DoF). The recovered trajectories for the case of the KLT features are shown in Fig. 7. The trajectory recovered using the planar motion model gave the results the most similar to the ground truth (Fig. 8). At this point, the reader might be wondering why the planar motion model performs still better than the circular motion alone. The reason is that the circular motion model is just an approximation of the real motion of the camera. Nevertheless, we can state that our restrictive model is very appropriate to describe and predict the motion of

the vehicle locally. Furthermore, its small parameterization (only 1 point) allows us to cope with those situations where only very few features are present, which usually cause other motion estimation algorithms to fail. This is also the reason why we put a firewall condition in our algorithm (Section V, step 7).

The comparison among the trajectories recovered using Harris, SIFT, and KLT is shown in Fig. 9. Here we show only the results for the planar motion model. Again, the paths exhibit different amounts of drift. After a deep inspection we found that the difference was not due to the quality of the feature matching (in fact we had always a reprojection error smaller than 0.5 pixels) but rather to the distribution of the features on each image. For instance we observed that while the KLT features were in many frames evenly distributed, the Harris features were sometimes more densely concentrated in some parts of the image. With SIFT the distribution of the features was more dramatic; observe, for instance, that around the turn pointed to by the arrow the path is more bent than the others. We found that this was due to the low number of SIFT features available (2 in this situation). This also happened for instance in scenes with very small clutter, like on bridges or in other areas where only the road, the lampposts, and the other vehicles were visible. In these situations the amount of SIFT features was much lower than Harris and KLT.

Figure 10 shows the top view of the recovered 3D map and camera trajectory. Here we used again the KLT features and the planar motion model. Observe that to build this map the features were triangulated only from two consecutive frames (two-view geometry). Furthermore observe that the points are aligned quite well along straight edges, which correspond to the walls of the buildings. Finally, Fig. 11 shows a closer view of the 3D map at the beginning of the path overlaid on a satellite image. Here it is more clear that the 3D points are well aligned along the straight edges of the buildings.

As mentioned earlier, the best results, in terms of agreement with the ground truth, were obtained using the KLT features and the planar motion model. The comparison is shown in Fig. 8. As observed the path is aligned quite well with the real trajectory except for the unavoidable drift that increases with the traveled distance. This result is however already quite good if one considers that the proposed approach is incremental (at each new frame only the current pose is updated without refining the previous ones) and that the position of the triangulated features is unchanged. Furthermore, the length of the recovered path was considerably long, up to 3Km. Improvements could be obtained by using for instance the recovered 3D structure to update the new pose (this would allow us to compute the relative scale from the images) or using bundle adjustment and SLAM techniques, where both the positions of the camera and the features are refined. These improvements are currently under development.
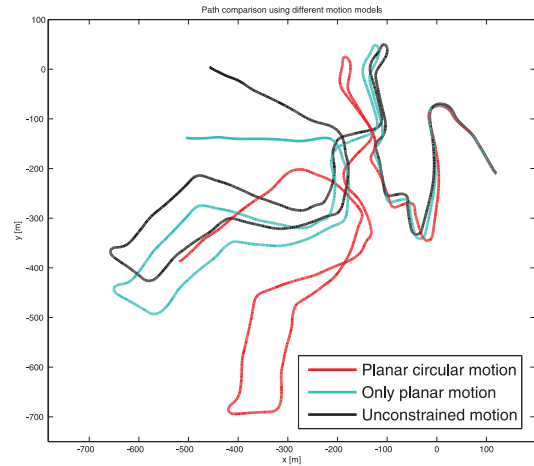

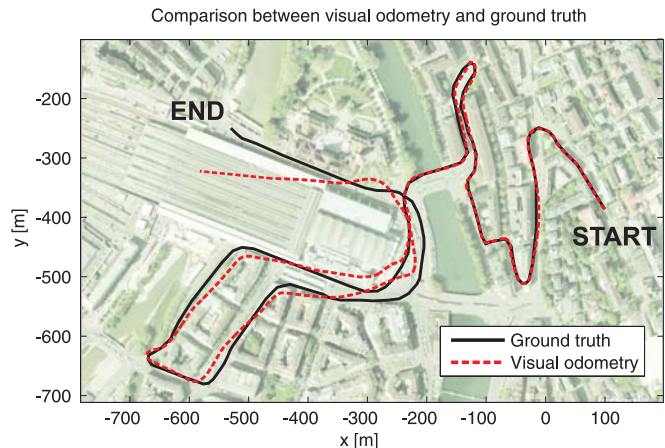Fig. 7. Visual odometry: comparison using different motion models. KLT features were used.

Fig. 8. Comparison between visual odometry (red dashed line) and ground truth (black solid line). The entire trajectory is 3Km long. For visual odometry here we used the KLT features and the planar motion model.

VII. CONCLUSION

In this paper, we showed that by exploiting the nonholonomic constraints of a wheeled vehicle it is possible to parameterize the motion with only 1 feature correspondence. Using a single feature correspondence for motion estimation is the lowest model parameterization possible and allowed us to develop the two most efficient algorithms for removing outliers: one based on RANSAC and the other one based on histogram voting. Furthermore, we showed that our restrictive motion model alone gives already accurate visual odometry estimates that can be refined from all the inliers using different motion models. In addition, we saw that this restrictive parameterization allows us to cope with those situations where only very few features are present, which usually cause other motion estimation algorithms to fail. The algorithm was tested on different feature detectors and we showed the performance of the approach by recovering a 3Km trajectory using omnidirectional images taken from our vehicle in a urban and very cluttered environment.
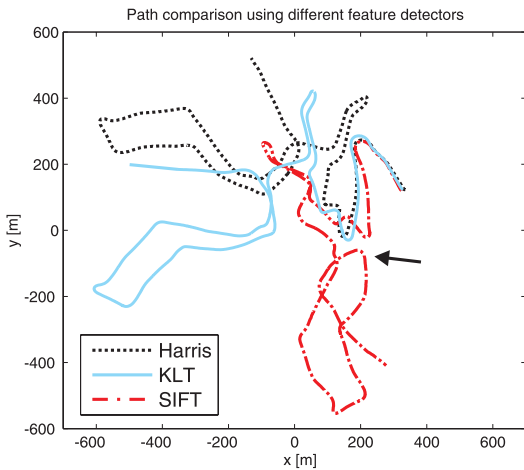
Fig. 9. Visual odometry: comparison using different feature detectors. Here, the planar motion model was used.
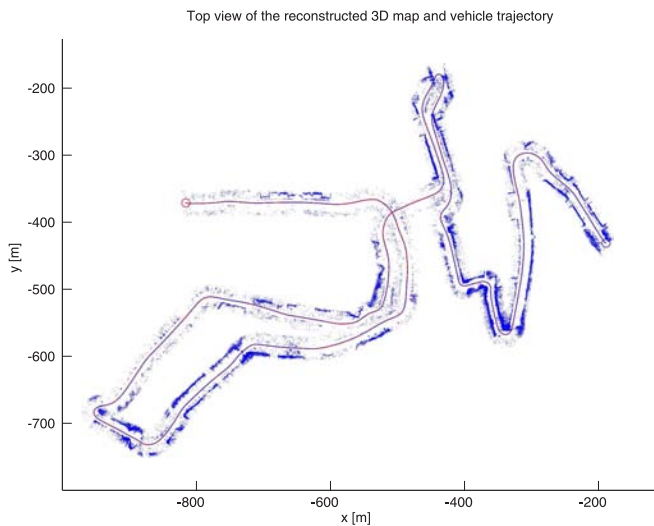

Fig. 10. Recovered 3D map and camera positions: top view. Here we used the KLT features and the planar motion model.
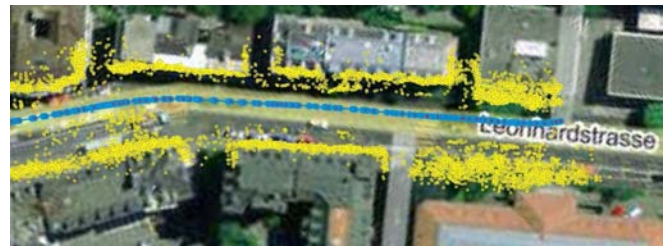

Fig. 11. A close-up of the recovered 3D map (yellow) overlaid on a satellite image. Here, we used KLT features. The camera positions are in blue. This image represents the street Leonhardstrasse in Zurich. The yellow cluttered points at the beginning of the path on the right represent trees.

REFERENCES

[1] Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Second edn. Cambridge University Press, ISBN: 0521540518 (2004)
[2] Bosse, M., Rikoski, R., Leonard, J., Teller, S.: Vanishing points and 3d lines from omnidirectional video. In: ICIP02. (2002) III: 513–516
[3] Corke, P.I., Strelow, D., , Singh, S.: Omnidirectional visual odometry for a planetary rover. In: IROS. (2004)
[4] Lhuillier, M.: Automatic structure and motion using a catadioptric camera. In: IEEE Workshop on Omnidirectional Vision. (2005)
[5] Nister, D., Naroditsky, O., , J., B.: Visual odometry for ground vehicle applications. Journal of Field Robotics (2006)
[6] Goecke, R., Asthana, A., Pettersson, N., Petersson, L.: Visual vehicle egomotion estimation using the fourier-mellin transform. In: IEEE Intelligent Vehicles Symposium. (2007)
[7] Tardif, J., Pavlidis, Y., Daniilidis, K.: Monocular visual odometry in urban environments using an omnidirectional camera. In: IEEE IROS'08. (2008)
[8] Milford, M.J., Wyeth, G.: Single camera vision-only slam on a suburban road network. In: IEEE International Conference on Robotics and Automation, ICRA'08. (2008)
[9] Scaramuzza, D., Siegwart, R.: Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. IEEE Transactions on Robotics, Special Issue on Visual SLAM **24** (2008)
[10] Nister, D.: Preemptive ransac for live structure and motion estimation. Machine Vision and Applications **16** (2005) 321–329
[11] Milford, M., Wyeth, G., Prasser, D.: Ratslam: A hippocampal model for simultaneous localization and mapping. In: International Conference on Robotics and Automation, ICRA'04. (2004)
[12] Scaramuzza, D., Fraundorfer, F., Pollefeys, M., Siegwart, R.: Closing the loop in appearance-guided structure-from-motion for omnidirectional cameras. In: Eighth Workshop on Omnidirectional Vision (OMNIVIS08). (2008)
[13] Deans, M.C.: Bearing-Only Localization and Mapping. PhD thesis, Carnegie Mellon University (2002)
[14] Davison, A.: Real-time simultaneous localisation and mapping with a single camera. In: International Conference on Computer Vision. (2003)
[15] Clemente, L.A., Davison, A.J., Reid, I., Neira, J., Tardos, J.D.: Mapping large loops with a single hand-held camera. In: Robotics Science and Systems. (2007)
[16] Lemaire, T., Lacroix, S.: Slam with panoramic vision. Journal of Field Robotics **24** (2007) 91–111
[17] Siegwart, R., Nourbakhsh, I.: Introduction to Autonomous Mobile Robots. MIT Press (2004)
[18] Fischler, M.A., Bolles, R.C.: RANSAC random sampling concensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of ACM **26** (1981) 381–395
[19] Nistér, D.: An efficient solution to the five-point relative pose problem. In: CVPR03. (2003) II: 195–202
[20] Stewenius, H., Engels, C., Nister, D.: Recent developments on direct relative orientation. ISPRS Journal of Photogrammetry and Remote Sensing **60** (2006) 284–294
[21] Philip, J.: A non-iterative algorithm for determining all essential matrices corresponding to five point pairs. Photogrammetric Record **15** (1996) 589–599
[22] Ortín, D., Montiel, J.M.M.: Indoor robot motion based on monocular images. Robotica **19** (2001) 331–342
[23] Scaramuzza, D., Martinelli, A., Siegwart, R.: A toolbox for easy calibrating omnidirectional cameras. In: IEEE International Conference on Intelligent Robots and Systems (IROS 2006). (2006)
[24] Scaramuzza, D.: Ocamcalib toolbox: Omnidirectional camera calibration toolbox for matlab (2006) Google for "*ocamcalib*".