

Accurate Visual Odometry from a Rear Parking Camera

Steven Lovegrove, Andrew J. Davison and Javier Ibañez-Guzmán

Abstract—As an increasing number of automatic safety and navigation features are added to modern vehicles, the crucial job of providing real-time localisation is predominantly performed by a single sensor, GPS, despite its well-known failings, particularly in urban environments. Various attempts have been made to supplement GPS to improve localisation performance, but these usually require additional specialised and expensive sensors. Offering increased value to vehicle OEMs, we show that it is possible to use just the video stream from a rear parking camera to produce smooth and locally accurate visual odometry in real-time. We use an efficient whole image alignment approach based on ESM, taking account of both the difficulties and advantages of the fact that a parking camera views only the road surface directly behind a vehicle.

Visual odometry is complementary to GPS in offering localisation information at 30Hz which is smooth and highly accurate locally whilst GPS is coarse but offers absolute measurements. We demonstrate our system in a large scale experiment covering real urban driving. We also present real-time fusion of our visual estimation with automotive GPS to generate a commodity-cost localisation solution which is smooth, accurate and drift free in global coordinates.

I. INTRODUCTION

Driving assistance systems are an increasingly important part of modern vehicles and various sensors are now fitted as standard to improve safety and aid in navigation. These systems rely more and more on knowledge of the vehicle's real-time position and orientation. Vehicles of the near future may need to refer to a global map for advance warning of approaching hazards such as intersections and pedestrian crossings [3], or may want to add to that map itself to share information with other road users. Localisation estimates for current vehicles come predominantly from automotive GPS, but its well known limitations, particularly in urban settings, mean that it cannot be relied on for the local accuracy needed for this type of function.



Fig. 1. A rear parking camera, as fitted to many current vehicles, views the road surface directly behind the car during normal driving.

In this paper we show that it is possible to use the video stream from a standard rear parking camera (Figure 1) to pro-

S. Lovegrove and A. J. Davison are with the Department of Computing, Imperial College London, UK.

J. Ibañez-Guzmán is with the Research Division, Renault S.A.S, 1 Av. du Golf, 78288 Guyancourt, France.

This research was supported by Renault, an EPSRC DTA studentship to S. Lovegrove and the European Research Council Starting Grant 210346.

duce a real-time trajectory estimate with local accuracy and smoothness which is highly complementary to automotive GPS. This camera, a standard feature of many current models, views the road surface directly behind the vehicle during normal driving. This surface (Figure 2) consists of mostly high frequency, self-similar, speckly texture. Although many point features can be detected in these images, they are individually highly ambiguous. Matching each part of the image successfully requires the support of whole frame.



Fig. 2. Four video frame samples taken from a parking camera in our test sequence.

Rather than feature matching, we use instead an efficient whole image alignment approach, based on the ESM algorithm and explicitly taking advantage of the near-planarity of the local road surface. Thereby we are able to make use of *all* of the texture present to robustly measure the image warp from one video frame to the next. This in turn leads to an accurate estimate of the vehicle's motion over the ground plane. A large advantage of the ESM method is that it is inherently parallelisable. Processing VGA images, we have a CPU only implementation that operates at 30 fps on an Intel i7 920 and a GPU accelerated implementation that achieves in excess of 300 fps with a single NVidia GTX 480 GPU.

In large scale urban driving experiments with a full ground-truth comparison, we show that our trajectory estimates from pure visual odometry are locally very accurate and smooth, though subject to inevitable drift over the long distances travelled in our full experiment. We go on to perform a live filtered fusion between our parking camera visual odometry estimates and an automotive GPS signal, and show that this combination of commodity level sensing available on many standard vehicles gives a quality of trajectory estimate comparable to an expensive PHINS system in robustness and accuracy.

II. RELATED WORK

In the motor industry, there has been much interest in using cameras to add functions related to safety and autonomous driving, with a particular emphasis on forward-looking stereo camera rigs to detect and estimate the locations of pedestrians or other vehicles. There have recently been attempts to use these cameras for estimation of the ego-motion of the vehicle itself. These approaches (e.g. [5], [2]) have mainly

followed the methods for visual odometry developed in the computer vision and robotics literature over recent years, based on point feature matching and geometry estimation over a sliding window. The first convincing stereo visual odometry system of this type was due to Nistér *et al.* [15], and state of the art performance is well represented by the recent work of Mei *et al.* [12]. Napier *et al.* tackled drift in their forward-looking stereo system by aligning frames to overhead satellite images [14].

General monocular visual odometry systems still lag somewhat in performance behind stereo systems, particularly due to the extra difficulty in reliably tracking enough high quality features to constrain motion, but good results can be achieved from video streams with high texture levels (e.g. [16]). The video stream available from a commodity vehicle parking camera is clearly not amenable to a general feature-based approach, due to the lack of distinctive texture in most frames. However, here we have the strong advantage of the assumptions that can be made about the constrained planar scene shape and vehicle motion, and this motivates our choice of a whole image alignment method. The ESM [10] algorithm we use is a development of Lucas and Kanade’s well known original alignment approach [9], and we are particularly motivated by its recent highly effective use in real-time spherical image mosaicing [8].

To our knowledge, we present the first convincing results for trajectory estimation based on a parking camera which views only the road surface. Azuma *et al.* [1] recently presented a method for estimating the motion of a car using a single forward-looking camera, which combined a standard feature-based approach in the top half of the image viewing street-side objects with an approach in principle similar to our own using homography estimation for the lower part of the image observing the road. However, many parts of their method were ad-hoc, and the results ultimately presented were limited (only one corner of real road). This was presumably because the robustness of the method was low. Our large scale results demonstrating both high accuracy and robustness go much further in proving the validity of estimating vehicle motion from road texture alignment, and have particular value due to the low-cost, software-only solution permitted by the use of an existing parking camera.

III. VISUAL ODOMETRY

In this paper, we choose to pose inter-frame odometry as an iterative continuous optimisation problem. For two consecutive video frames, the image warp that transforms the visible portion of one frame to the other is uniquely described by the video camera’s motion and the structure of the scene being observed. For a rear mounted parking camera, video frames recorded whilst driving will largely image the ground upon which the vehicle is moving. We treat the ground as a smooth, locally planar manifold, and produce odometry relative to that manifold.

Our simple model assumes that the car moves perfectly parallel to the ground and inter-frame vehicle motion can be parametrised by the vector $(\Delta x, \Delta y, \Delta \theta)^T$. The constrained

dynamics of a vehicle allow us to model motion with a constant velocity prior. This helps us in our optimisation to find the true minima over other local minima which may exist.

A. Parametrisation

The camera’s motion is a function of the vehicle’s motion and thus we consider two important frames of reference, that of the camera and that of the car (Figure 3).

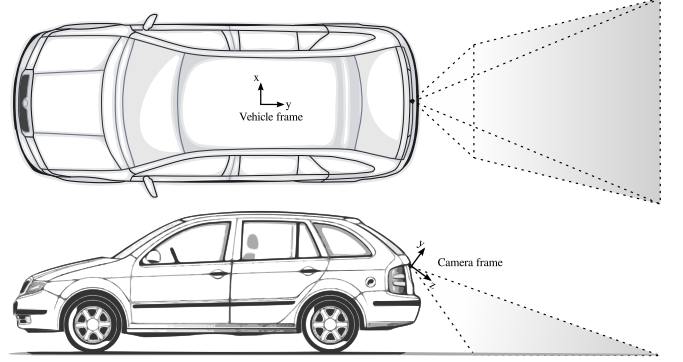


Fig. 3. Calibrated car centric and camera centric frames of reference.

We describe the two-dimensional location of the vehicle with respect to the world whilst taking image \mathcal{I}^i by T^{wv_i} , a homogeneous point transfer matrix belonging to the Lie Group of rigid body transforms in two-dimensions, $SE(2)$. For two images \mathcal{I}^l and \mathcal{I}^r with associated positions T^{wv_l} and T^{wv_r} , we can calculate the vehicle transformation between them by matrix composition, $T^{v_l v_r} = (T^{wv_l})^{-1} T^{wv_r} = T^{v_l w} T^{w v_r}$.

Since the parking camera captures images of the world projectively in three-dimensions, we also describe the pose of the camera whilst imaging frame \mathcal{I}^i , by T^{wc_i} , a homogeneous point transfer matrix belonging to the Lie Group of rigid body transforms in three-dimensions, $SE(3)$. Although T^{wv_i} is a two-dimensional transform, we can arbitrarily refer to it in three-dimensions by ‘raising’ it into $SE(3)$ by enforcing that the vehicle lives on the X-Y plane. We do so implicitly for notational convenience.

As the camera is fixed rigidly to the vehicle, there is a constant transform $T^{vc} = (T^{wv_i})^{-1} T^{wc_i}$ which relates the vehicle frame of reference to that of the camera, for all i . This allows us to write the three-dimensional transformation of the camera between images \mathcal{I}^r and \mathcal{I}^l as a function of the vehicle motion.

$$T^{c_l c_r} = (T^{vc})^{-1} T^{v_l v_r} T^{vc} \quad (1)$$

B. Cost Function

Observations of a plane via a projective camera are related to one another via a plane induced homography:

$$H^{lr} = K T^{c_l c_r} (I - \mathbf{n}_{d_c})^T K^{-1}. \quad (2)$$

Here, $H^{lr} \in \mathbb{R}^{3 \times 3}$ describes the homogeneous transform that takes pixels in the ‘reference’ camera image, \mathcal{I}^r , to

those in the ‘live’ image, \mathcal{I}^l . $K \in \mathbb{R}^{3 \times 3}$ is referred to as the intrinsics matrix which projects points in three-dimensions into the camera. $\mathbf{n}_{d_c} = \frac{\hat{\mathbf{n}}_c}{d_c}$ describes the pose and height of the camera relative to the ground, where $\hat{\mathbf{n}}_c$ is the unit normal of the plane in the camera frame of reference c and d_c is the distance of closest approach to the plane.

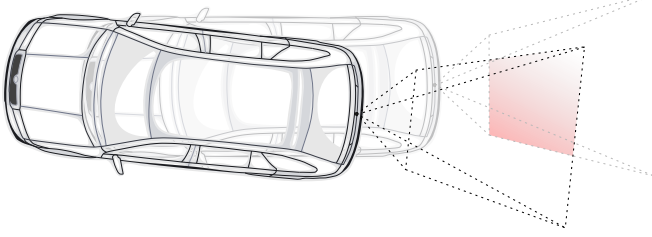


Fig. 4. A co-visible region (shaded red) exists between consecutive camera images. A linear mapping described by a plane induced homography lets us transfer pixels from one image to another.

Between two consecutive image frames, \mathcal{I}^l and \mathcal{I}^r (Figure 4), if we were to know the parameters of K , T^{vc} , T^{vlv_r} and \mathbf{n}_{d_c} , we could synthesize one frame from the other by transferring the pixels via H^{lr} (Equation 2). For any given synthesis, we can compare how similar it is to the true image by comparing how similar their pixels are. Assuming that the parking camera is fixed rigidly to the vehicle, and following from the assumption that it is travelling along a locally planar surface, K , T^{vc} and \mathbf{n}_{d_c} will not change. They can be easily estimated via off the shelf camera calibration packages. This leaves the motion of the vehicle T^{vlv_r} unknown, and is in fact the quantity we wish to estimate.

Instead of measuring T^{vlv_r} directly, we progress by saying that we have an estimate of the vehicle motion \hat{T}^{vlv_r} (which may come from a motion model such as a constant velocity prior) and wish to find some refinement parametrised by some vector \mathbf{x} . With this in mind, we define a cost function, $F(\mathbf{x})$, which is equal to the sum of squared pixel error between a synthetic image generated by warping \mathcal{I}^l by $H^{lr}(\mathbf{x})$ and the true image, \mathcal{I}^r . The sum is taken for every pixel p_r in image \mathcal{I}^r , belonging to the region Ω_r . Per pixel differences $f_{p_r}(\mathbf{x})$ can be written as a stacked column vector $f(\mathbf{x})$:

$$F(\mathbf{x}) = \frac{1}{2} \sum_{p_r \in \Omega_r} \left(f_{p_r}(\mathbf{x}) \right)^2 = \frac{1}{2} \|f(\mathbf{x})\|^2, \quad (3)$$

$$f_{p_r}(\mathbf{x}) = \mathcal{I}^l \left(\pi \left(H^{lr}(\mathbf{x}) p_r \right) \right) - \mathcal{I}^r \left(\pi(p_r) \right), \quad (4)$$

where $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is the lowering function which performs homogeneous division. We rewrite Equation 2 using Equation 1 and \hat{T}^{vlv_r} as:

$$H^{lr}(\mathbf{x}) = K T^{cv} \hat{T}^{vlv_r} T(\mathbf{x}) T^{vc} (I - \mathbf{n}_{d_c})^T K^{-1}. \quad (5)$$

$T(\mathbf{x})$ represents a small change to the estimate \hat{T}^{vlv_r} . We wish to find the vector $\mathbf{x} = \mathbf{x}_0$ which minimises $F(\mathbf{x})$, so we would like to do so iteratively, estimating \mathbf{x}_0 by $\hat{\mathbf{x}}$ and applying the update rule:

$$\hat{T}^{vlv_r} \leftarrow \hat{T}^{vlv_r} T(\hat{\mathbf{x}}). \quad (6)$$

This process is repeated until convergence which may be detected based on a threshold on the magnitude of the update $\hat{\mathbf{x}}$. We choose to parametrise our update matrix $T(\mathbf{x})$ minimally by $\mathbf{x} \in \mathbb{R}^3$ belonging to $\mathfrak{se}(2)$, the Lie Algebra of the Lie Group $SE(2)$. This parametrisation is locally euclidean around the identity transform and for $SE(2)$ is equivalent to the vector $(\Delta x, \Delta y, \Delta \theta)^T$. The Lie Algebra $\mathfrak{se}(2)$ is related to the Lie Group $SE(2)$ by:

$$T(\mathbf{x}) = \exp \left(\sum_{i=1}^3 \mathbf{x}_i \mathbf{A}_i \right), \quad (7)$$

where $\mathbf{A}_{0,1,2} \in \mathbb{R}^{3 \times 3}$ are the group generators for $SE(2)$ [6].

One solution to finding \mathbf{x}_0 and therefore the inter-frame motion of the vehicle would be to sample from the function $F(\mathbf{x})$ to find its local minimiser. For a live system this is clearly undesirable in terms of computational cost and it would be difficult to decide on a sufficient sampling resolution in the space of \mathbf{x} .

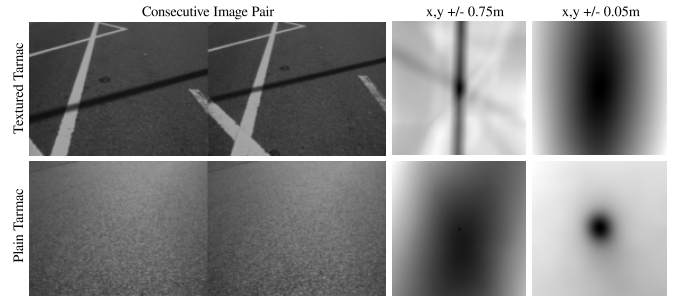


Fig. 5. Two sample image pairs from sequence (top left, bottom left) with visualisation of cost landscape, centred around the chosen minima and normalised to lowest cost (black) and highest visible cost (white), for +/- 0.75 m (middle) and +/- 0.05m (right).

Figure 5 shows the cost landscape for varying the first two ordinates of \mathbf{x} (Δx and Δy) for two separate consecutive image pairs in our test sequence. The cost landscape plots are centred around the true minima and are intensity normalised to display lowest sampled cost as black, and highest sampled cost as white. For the more structured image pair containing road markings, we can see that the horizontal direction is more heavily constrained than the forward-backward Δy direction. This is caused by the strong road markings which lie largely in the direction of travel. The more structured texture in this pair creates an extremely steep horizontal cost basin.

For the plain tarmac, we see a much flatter cost landscape, with a very sharp basin containing the minima. This reflects the uniform self similarity of the images and motivates a technique which considers the images as a whole.

C. Efficient Second order Minimisation

In order to approximate \mathbf{x}_0 closely but efficiently, we follow the method proposed by Malis of Efficient Second order Minimisation (ESM) [10] with further contributions

from Mei *et al.* [11]. ESM descends on the solution iteratively, taking second-order steps within the cost space whilst computing only first-order terms. For completeness, this section will derive the ESM update for a general sum of squared error cost function. The development and convergence / performance analysis of ESM compared to other prevalent methods can be found within [10].

Defining $\nabla f(\mathbf{x})$ to be the row vector of partial derivatives of f , evaluated at \mathbf{x} , we start by taking the Taylor series expansion of both $f(\mathbf{x})$ and $\nabla f(\mathbf{x})$ about $\mathbf{0}$:

$$f(\mathbf{x}) = f(\mathbf{0}) + \nabla f(\mathbf{0})\mathbf{x} + \frac{1}{2}\mathbf{x}^T \nabla \nabla f(\mathbf{0})\mathbf{x} + \dots \quad (8)$$

$$\nabla f(\mathbf{x}) = \nabla f(\mathbf{0}) + \mathbf{x}^T \nabla \nabla f(\mathbf{0}) + \dots \quad (9)$$

Equating $\mathbf{x}^T \nabla \nabla f(\mathbf{0})$ in Equations 8 and 9, we can approximate $F(\mathbf{x})$ up to second order as a function of first order terms, $\hat{F}(\mathbf{x}) \approx F(\mathbf{x})$, $\hat{f}(\mathbf{x}) \approx f(\mathbf{x})$:

$$\hat{f}(\mathbf{x}) = f(\mathbf{0}) + \frac{1}{2}(\nabla f(\mathbf{0}) + \nabla f(\mathbf{x}))\mathbf{x} \quad (10)$$

$$\hat{F}(\mathbf{x}) = \frac{1}{2}\|\hat{f}(\mathbf{x})\|^2 = \frac{1}{2}\hat{f}(\mathbf{x})^T \hat{f}(\mathbf{x}). \quad (11)$$

A simple application of the product rule allows us to approximate $\nabla F(\mathbf{x})$ from Equation 11. By definition, the local minimum of $F(\mathbf{x})$ represents a stationary point of that function (Equation 13), and we make the assumption that such a minimum exists.

$$\nabla F(\mathbf{x}) \approx \nabla \hat{f}(\mathbf{x})^T \hat{f}(\mathbf{x}). \quad (12)$$

$$\nabla F(\mathbf{x}_0) = \mathbf{0}. \quad (13)$$

It follows from Equations 10, 12 and 13 that we can find the approximate local minimiser of $F(\mathbf{x})$ by solving:

$$\nabla \hat{f}(\mathbf{x}_0)^T \left(f(\mathbf{0}) + \frac{1}{2}(\nabla f(\mathbf{0}) + \nabla f(\mathbf{x}_0))\mathbf{x}_0 \right) = \mathbf{0} \quad (14)$$

$$\mathbf{J}\mathbf{x}_0 = -f(\mathbf{0}) \quad (15)$$

Writing $\mathbf{J} = \frac{1}{2}(\nabla f(\mathbf{0}) + \nabla f(\mathbf{x}_0))$, solving Equation 14 is equivalent to solving Equation 15. We can instead solve this overdetermined system in the linear least squares sense by considering its normal equations:

$$\mathbf{J}^T \mathbf{J} \mathbf{x}_0 = -\mathbf{J}^T f(\mathbf{0}) \quad (16)$$

$$\mathbf{x}_0 = -(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T f(\mathbf{0}) \quad (17)$$

We see that in order to find the solution \mathbf{x}_0 in Equation 17, we are required to evaluate the partial derivatives of the cost function at $\mathbf{0}$ and at the solution \mathbf{x}_0 . At first this might seem difficult, but returning to our specific cost function (Equation 4), we can express its partial derivatives per pixel, through the chain and product rules, as follows:

$$\begin{aligned} \frac{\partial f_{pr}(\mathbf{x})}{\partial \mathbf{x}_i} &= \frac{\partial T^l(\mathbf{a})}{\partial \mathbf{a}} \bigg|_{\mathbf{a}=\pi(\mathbf{H}^{lr}(\mathbf{x})p_r)} \cdot \frac{\partial \pi(\mathbf{b})}{\partial \mathbf{b}} \bigg|_{\mathbf{b}=\mathbf{H}^{lr}(\mathbf{x})p_r} \cdot \frac{\partial \mathbf{H}^{lr}(\mathbf{x})}{\partial \mathbf{x}_i} p_r \\ \frac{\partial \mathbf{H}^{lr}(\mathbf{x})}{\partial \mathbf{x}_i} &= \mathbf{K} \mathbf{T}^{cv} \hat{\mathbf{T}}^{v_l v_r} \frac{\partial \mathbf{T}(\mathbf{x})}{\partial \mathbf{x}_i} \mathbf{T}^{vc} (\mathbf{I} - \mathbf{n}_{d_c})^T \mathbf{K}^{-1} \end{aligned} \quad (18)$$

Since the Lie Algebra $\mathfrak{se}(2)$ defines a locally invariant vector field tangential to the manifold of $SE(2)$, $\frac{\partial \mathbf{T}(\mathbf{x}_0)}{\partial \mathbf{x}_i} \mathbf{x}_0 = \frac{\partial \mathbf{T}(\mathbf{0})}{\partial \mathbf{x}_i} \mathbf{x}_0$ for small \mathbf{x}_0 . This can be shown to render the terms $\frac{\partial \mathbf{H}^{lr}(\mathbf{0})}{\partial \mathbf{x}_i}$ and $\frac{\partial \mathbf{H}^{lr}(\mathbf{x}_0)}{\partial \mathbf{x}_i}$ equivalent and they are constant for each pixel within the iteration. $\frac{\partial \mathbf{T}(\mathbf{0})}{\partial \mathbf{x}_i} = \mathbf{A}_i$, the i th group generator for $SE(2)$ as used in Equation 7.

Finally, in order to evaluate the image gradient in the live image for the solution vector \mathbf{x}_0 , the ESM method draws on the fact that at the true minimiser, the two images are directly aligned, so we take the image gradient of the reference image instead.

IV. EXPERIMENTAL SETUP

Our evaluation is against data obtained during an extended experiment where a multipurpose Renault Espace passenger vehicle conducted a 2.5km run through an urban setting, capturing images continuously at 30fps while travelling at speeds of up to 45km/h. Data from a U-blox automotive type GPS receiver is used during our data fusion experiments. The rear view camera is a Fire-i digital camera having VGA resolution.

On board was also a ground truth system capable of estimating with high-accuracy the vehicle's position and orientation. This system consists of a tactical level Inertial Measurement Unit made of 3 fibre optic gyroscopes and 3 pendulum type accelerometers, a bi-frequency GPS receiver and the vehicle wheel odometry [4]. Despite occlusion of GPS signals or multipaths, the ground truth estimations remain precise with positioning error below half a meter.

The intrinsic parameters \mathbf{K} of the parking camera used were calibrated using a standard calibration grid technique. The camera's extrinsic location relative to the vehicle frame \mathbf{T}^{cv} was initially hand-measured, but this estimate was then refined by visual means in a process similar to that developed by Miksch *et al.* [13].

V. RESULTS

A. Drifting Odometry

Our visual odometry system estimates inter-frame motion at frame rate (30Hz). This measurement is most useful to determine the velocity of the vehicle. Figure 6 shows the linear and angular velocity as measured by our system compared to that of the ground truth PHINS system on a five and a half minute road sequence. Visual odometry is plotted at 30Hz whereas the PHINS system is plotted at 1Hz. The bottom plot shows the system's confidence as measured by the sum of squared pixel error between aligned frames. We can see that the visual odometry measurement follows the ground truth system closely. There are four short sections in

the sequence where tracking could not provide an estimate and here constant velocity is assumed (Annotation 2).

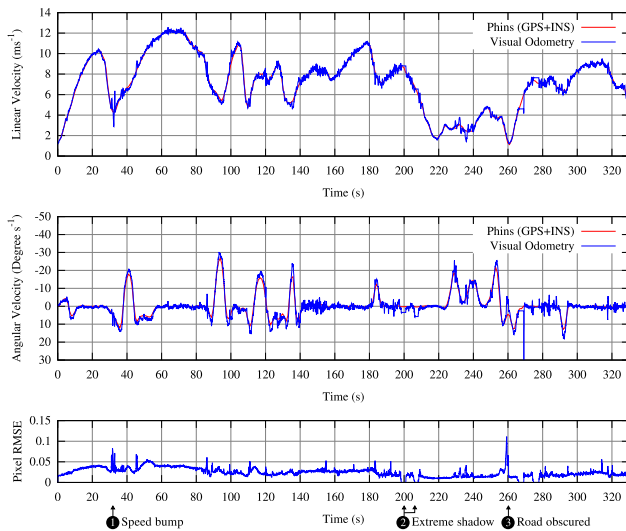


Fig. 6. Linear and angular velocity of visual odometry compared with ground truth over time for 2.5km sample sequence.

We can integrate our measured velocity over time to produce a trajectory in some common frame of reference. This, like *any* form of motion estimation based only on relative local sensing such as vision, will of course drift away from the frame of reference over time. Figure 7 shows stretches of trajectory estimate, each one minute long, aligned in each case to ground truth at the first frame. This helps us to visualise the rate at which the odometry drifts from an absolute map.



Fig. 7. Four minute-long sequences of integrated visual odometry (blue) against ground truth (green).

Figure 8 shows four ten second stretches of trajectory, again each aligned on the first frame to ground truth. We

additionally show standard automotive GPS overlaid for comparison. We can see that our visual odometry system is much smoother locally than the GPS measurements and is sampled at a higher frequency.



Fig. 8. Four 10 second sequences of integrated visual odometry (blue) against GPS (red) and ground truth (green).

B. Visual Odometry fused with GPS

The benefits of visual odometry and GPS are clearly complimentary. Visual odometry offers a robust, accurate and smooth local estimate whereas GPS provides coarse measurements fixed to a global map. We performed a real-time sliding window optimisation which incorporated both visual odometry and GPS to provide accurate drift-free vehicle localisation. We used g2o, a general framework for graph optimisation which allowed us to enter relative edges corresponding to inter-frame measurements, and unary edges corresponding to absolute GPS measurements [7].

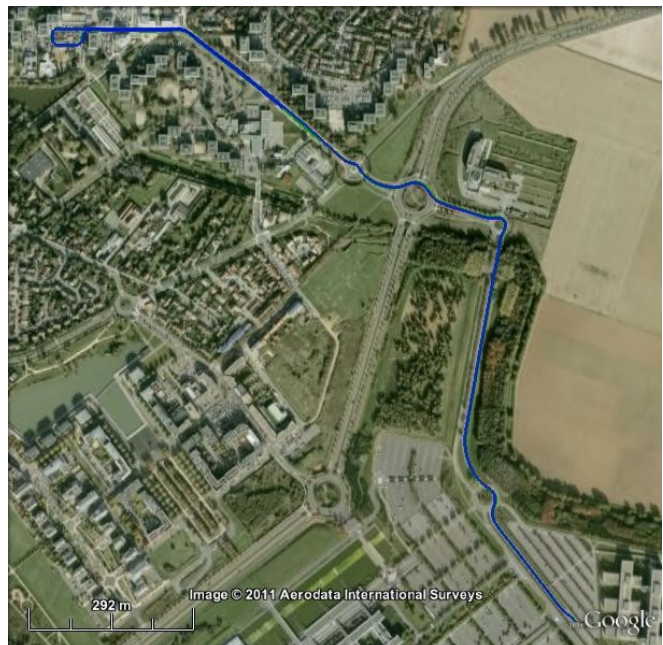


Fig. 9. Fused visual odometry and GPS (blue) against ground truth (green).

Figure 9 shows the resulting trajectory after sensor fusion. We can see that the resulting system deviates only a small

amount from the ground truth system and it is hard at this scale to differentiate them. Figure 10 shows an enlargement which includes also the automotive GPS trajectory. We can see that compared to GPS, the combination of GPS and visual odometry is much smoother, whilst accurately capturing the path of the vehicle through the roundabout.

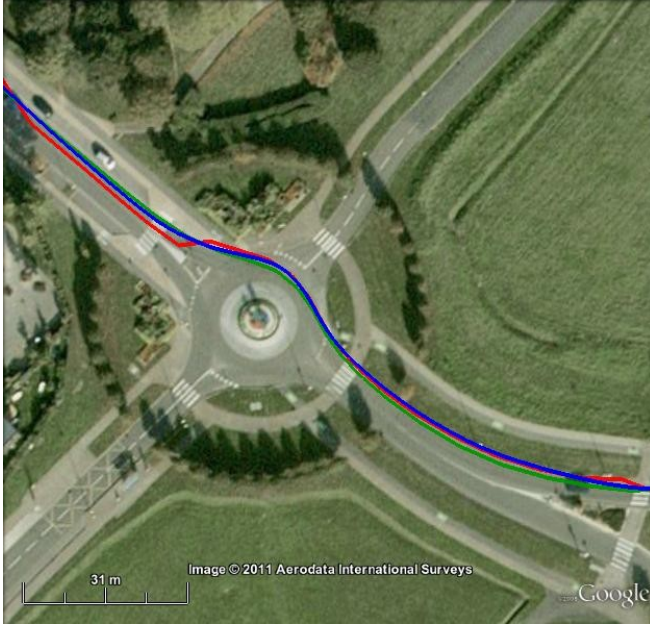


Fig. 10. Close up sample of fused visual odometry and GPS (blue) against GPS only (red) and ground truth (green).

C. Failure Modes

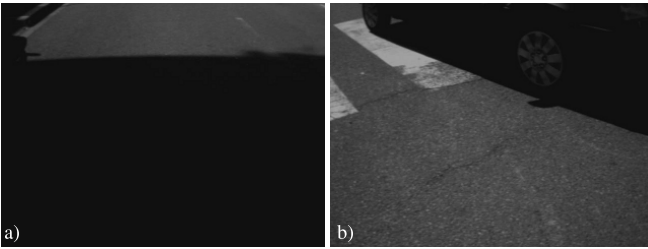


Fig. 11. a) The vehicle passes under heavy shadow causing the image frame to become completely saturated by pure black. b) Another vehicle obscures a large amount of the road, and induces incorrect motion.

There are two main failure modes in our system (Figure 11). The first comes from lighting which exceeds the dynamic range of the camera; there are four short sections in the sequence which are also identifiable in Figure 6 where the video frames are completely saturated black. Clearly, no software solution could do any better than relying on motion priors or extra sensors such as wheel odometry. Such a situation is clearly detectable by measuring the image variance. We assume constant velocity motion when image variance passes below a set-once threshold.

The second failure mode comes when our basic scene assumption is violated. Annotation 1 in Figure 6 shows

the velocity measurement error induced by a non-planar speed bump and the associated increase in mean squared pixel error. Annotation 3 highlights this assumption being broken again when the road is part obscured by a turning vehicle (Figure 11b). These situations can be considered during sensor fusion by making use of the visual odometry's confidence output, mitigating their overall influence.

VI. CONCLUSION

We have presented a robust method for estimating accurate vehicle odometry from images obtained from a rear-facing parking camera. We have demonstrated how this might be combined with standard automotive GPS to provide accurate 30Hz localisation of a vehicle relative to a global map and suggest that such localisation is a clear benefit to future autonomous and safety systems. The failure cases of the system are infrequent and when they do occur, they can be detected.

REFERENCES

- [1] T. Azuma, S. Sugimoto, and M. Okutomi. Egomotion estimation using planar and non-planar constraints. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2010.
- [2] A. Bak, S. Bouchafa, and D. Aubert. Detection of independently moving objects through stereo vision and ego-motion estimation. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2010.
- [3] J. Ibanez-Guzmán, S. Lefevre, and A. K. Mokaddem. Vehicle to vehicle communications applied to road intersection safety. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2010.
- [4] IXSEA, Landins: Georeferencing and positioning system. URL <http://www.ixsea.com/en/products/10/landins.html>, 2010.
- [5] B. Kitt, A. Geiger, and H. Lategahn. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2010.
- [6] G. Klein. *Visual Tracking for Augmented Reality*. PhD thesis, University of Cambridge, 2006.
- [7] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g^2o : A general framework for graph optimization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- [8] S. Lovegrove and A. J. Davison. Real-time spherical mosaicing using whole image alignment. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2010.
- [9] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 1981.
- [10] E. Malis. Improving vision-based control using efficient second-order minimization techniques. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2004.
- [11] C. Mei, S. Benhimane, E. Malis, and P. Rives. Efficient homography-based tracking and 3-D reconstruction for single-viewpoint sensors. *IEEE Transactions on Robotics (T-RO)*, 24(6):1352–1364, 2008.
- [12] C. Mei, G. Sibley, M. Cummins, P. Newman, and I. Reid. A constant time efficient stereo SLAM system. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2009.
- [13] M. Miksch, B. Yang, and K. Zimmerman. Automatic extrinsic camera self-calibration based on homography and epipolar geometry. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2010.
- [14] A. Napier, G. Sibley, and P. Newman. Real-time bounded-error pose estimation for road vehicles using vision. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2010.
- [15] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [16] H. Strasdat, J. M. M. Montiel, and A. J. Davison. Scale drift-aware large scale monocular SLAM. In *Proceedings of Robotics: Science and Systems (RSS)*, 2010.