6D Visual Odometry with Dense Probabilistic Egomotion Estimation

Hugo Silva[1], Alexandre Bernardino [2] and Eduardo Silva[1]
[1]*INESC TEC Robotics Unit, ISEP, Rua Dr. Antonio Bernardino de Almeida 431, Porto, Portugal*
[2]*Institute of Systems and Robotics, IST, Avenida Rovisco Pais 1, Lisboa, Portugal*

Keywords: Visual Navigation, Stereo Vision, Visual Odometry, Egomotion.

Abstract: We present a novel approach to 6D visual odometry for vehicles with calibrated stereo cameras. A dense probabilistic egomotion (5D) method is combined with robust stereo feature based approaches and Extended Kalman Filtering (EKF) techniques to provide high quality estimates of vehicle's angular and linear velocities. Experimental results show that the proposed method compares favorably with state-the-art approaches, mainly in the estimation of the angular velocities, where significant improvements are achieved.

1 INTRODUCTION

Visual Odometry is the term generically used to denote the process of estimating linear and angular velocities of a vehicle equipped with vision cameras (Scaramuzza and Fraundorfer, 2011). These systems are becoming ubiquitous in mobile robotics applications due to the availability of low-cost high quality cameras and their ability to complement the measurements provided by Inertial Measurement Units (IMU). Because vision sensors ground their perception on static features of the environment, they are in principle less prone to the estimation bias rather common on IMU sensors. In this work we focus on the development of Visual Odometry Systems for mobile robots equipped with a calibrated stereo camera setup.

Visual Odometry Systems are an important component on mobile robot's navigation systems. The short term velocity estimates provided Visual Odometry has been shown to improve the localization results of Simultaneous Localization and Mapping (SLAM) methods. For instance in (Alcantarilla et al., 2010), Visual Odometry measurements are used as priors for the prediction step of a robust EKF-SLAM algorithm.

Visual Odometry systems have been continuously developed over the past 30 years. These systems suffered a major outbreak due to the outstanding work of (Maimone and Matthies, 2005) on NASA Mars Rover Program. Nister ((Nistér, 2004)) developed a Visual Odometry system, based on a 5-point algorithm, that became the standard algorithm for comparison of Visual Odometry techniques. This algorithm can be used either in stereo or monocular vision approaches and consists on the use of several visual processing techniques, namely: feature detection and matching, tracking, stereo triangulation and RANSAC (Fischler and Bolles, 1981) for pose estimation with iterative refinement.

In (Moreno et al., 2007) it is proposed a visual odometry estimation method using stereo cameras. A closed form solution is derived for the incremental movement of the cameras and combines distinctive features SIFT (Lowe, 2004) with sparse optical flow.

There are already some approaches to stereo visual odometry estimation using dense methods like the one developed by (Comport et al., 2007), that uses a quadrifocal warping function to track features using dense correspondences to correctly estimate 3D visual odometry.

In (Domke and Aloimonos, 2006), a method for estimating the epipolar geometry describing the motion of a camera is proposed using dense probabilistic methods. Instead of deterministically choosing matches between two images, a probability distribution is computed over all possible correspondences. By exploiting a larger amount of data, a better performance is achieved under noisy measurements. However, that method is more computationally expensive and does not recover translational scale factor.

In our work, we propose the use of a dense probabilistic method such as in (Domke and Aloimonos, 2006) but with two important additions: (i) a sparse feature based method is used to estimate the translational scale factor and (ii) a fast correspondence method using a recursive ZNCC implementation is provided for computational efficiency.

Silva H., Bernardino A. and Silva E..
6D Visual Odometry with Dense Probabilistic Egomotion Estimation.
DOI: 10.5220/0004196103610365
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 361-365
ISBN: 978-989-8565-48-8

Our method, denoted 6DP combines sparse feature detection and tracking for stereo-based depth estimation, using highly distinctive SIFT features (Lowe, 2004) and a variant of the dense probabilistic ego-motion method developed by (Domke and Aloimonos, 2006) to estimate camera motion up to a translational scale factor. Upon obtaining two registered point sets in consecutive time frames, an Absolute Orientation method, defined as an orthogonal Procrustes problem (AO) is used to recover yet undetermined motion scale. The velocities obtained by the proposed method are then filtered with a EKF approach to reduce sensor noise and provide frame-to-frame filtered linear and angular velocity estimates.

Our method was compared with the methods in LIBVISO Visual Odometry Library (Kitt et al., 2010), using standard dataset from this library. Ground truth is also provided, through the fusion of IMU and GPS measurements. Results show that our method presents significant improvements in the estimation of angular velocities and a similar performance for linear velocities. The benefits of using dense probabilistic approaches are thus validated in a real world scenario with practical significance.

2 6D VISUAL ODOMETRY USING DENSE AND SPARSE EGO-MOTION ESTIMATION

Our solution is based on the probabilistic method of egomotion estimation using the epipolar constraint developed by (Domke and Aloimonos, 2006). However, the method from (Domke and Aloimonos, 2006) is unable to estimate motion scale, so a stereo vision sparse feature based approach that uses detected SIFT features correspondence between I_{Tk} and I_{Tk+1} is used to obtain translation motion scale.

An architecture of our method is displayed in figure 1. In summary, it is composed by the following main steps:

1. First, SIFT feature points are detected in the current pair of stereo frames (I^L_{Tk}, I^R_{Tk}), using a known feature detector. These image feature points are then correlated between left and right image to obtain 3D point depth information.
2. Second, we use a dense image pixel correlation method, that due to its probabilistic nature, does not commit the match correlation of image point $P_k(x,y)$ in I^L_{Tk} to other image point $P_k(x,y)$ in I^L_{Tk+1}. Instead, it copes with several hypothesis of matching for image point $P_k(x,y)$ in I^L_{Tk+1}, thus making the estimation of the essential matrix

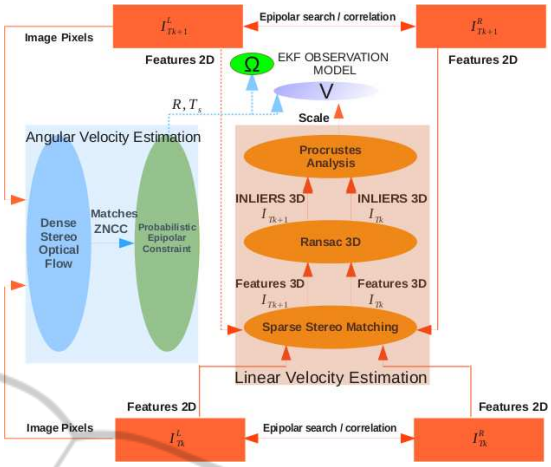

Figure 1: 6D Visual Odometry System Architecture.

E_s more robust to image feature matching errors and hence providing a more accurate camera motion estimation [R,t] between I_{Tk} and I_{Tk+1}. The dense likelihood correspondence maps are computed based on ZNCC(Huang et al., 2011) correlation.

3. Third, due to the need to determine the motion scale between I_{Tk} and I_{Tk+1}, a Procrustes absolute orientation method(AO) is utilized. The AO method uses 3D image feature points obtained by triangulation from stereo image pairs (I^L_{Tk}, I^R_{Tk}) and (I^L_{Tk+1}, I^R_{Tk+1}) combined with robust techniques like RANSAC(Fischler and Bolles, 1981), thus obtaining only good candidates (inliers) for Procrustes based motion scale determination.
4. Finally, vehicle linear and angular velocity (V, Ω) between I_{Tk} and I_{Tk+1} is determined.

All, of this steps are then encapsulated within an Extended Kalman filter yielding a more robust camera motion estimation.

2.1 Probabilistic Correspondence

The key to the proposed method relies in the consideration of probabilistic rather than deterministic matches. Usual methods for motion estimation consider a match function M that associates coordinates of points $\mathbf{m} = (x,y)$ in image 1 to points $\mathbf{m}' = (x',y')$ in image 2 :

$$M(\mathbf{m}) = \mathbf{m}' \tag{1}$$

Instead, the probabilistic correspondence method defines a probability distribution over the points in image 2 for all points in image 1:

$$P_{\mathbf{m}}(\mathbf{m}') = P(\mathbf{m}'|\mathbf{m}) \tag{2}$$

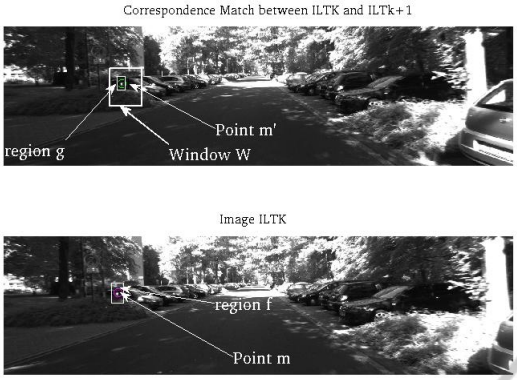

Figure 2: Image feature Point correspondence for ZNCC matching.

Thus, all points $\mathbf{m}'$ in image 2 are candidates for matching with point $\mathbf{m}$ in image 1 with *a priori* likelihoods proportional to $P_{\mathbf{m}}(\mathbf{m}')$. One can consider $P_{\mathbf{m}}$ as images (one per each pixel in image 1) whose value in $\mathbf{m}'$ is proportional to the likelihood of $\mathbf{m}'$ matching with $\mathbf{m}$. For the sake of computational cost, likelihoods are not computed for the whole range in image 2 but just to windows around $\mathbf{m}$ (or suitable predictions given prior information), see figure 2.

In (Domke and Aloimonos, 2006) this value was computed via the normalized product, over a filter bank of Gabor filters with different orientation and scales, of the exponential of the negative differences between the angle of the Gabor filter responses in $\mathbf{m}$ and $\mathbf{m}'$.

The motivation for using a Gabor filter bank was the robustness of their responses to changes in the brightness and contrast of the image. However, the computations demand a significant computational effort, thus we propose to perform the computations with the well known Zero Mean Normalized Cross Correlation function(ZNCC).

This function is also known to be robust to brightness and contrast changes and recent efficient recursive schemes developed by Huang et al (Huang et al., 2011) render it suitable to real-time implementations. That method is faster to compute and yields the same quality as the method of Domke.

2.1.1 Probabilistic Egomotion

From two images of the same camera, one can recover its motion up to the translation scale factor. This can be represented by the epipolar constraint which, in homogeneous normalized coordinates can be written as:

$$(\mathbf{s}')^T E\mathbf{s} = 0 \tag{3}$$

where E is the so called Essential Matrix (Hartley and Zisserman, 2004), a $3X3$ matrix with rank 2 and 5 degrees-of-freedom. Intuitively, this matrix expresses the directions in image 2 that should be searched for matches of points in image 1. It can be factored by:

$$E = R[t]_{\times} \tag{4}$$

where R and t are, respectively, the rotation and translation of the camera between the two frames.

To obtain the Essential matrix from the probabilistic correspondences, (Domke and Aloimonos, 2006) proposes the computation of a probability distribution over the (5-dimensional) space of essential matrices. Each dimension of the space is discretized in 10 bins, thus leading to 100000 hypotheses E_i. For each point $\mathbf{s}$ the likelihood of these hypotheses are evaluated by:

$$P(E_i|\mathbf{s}) \propto \overset{max}{\mathbf{s}'} : (\mathbf{s}')^T E_i \mathbf{s} = 0 \; P_{\mathbf{s}}(\mathbf{s}') \tag{5}$$

Intuitively, for a single point $\mathbf{s}$ in image 1, the likelihood of a motion hypothesis is proportional to the best match obtained along the epipolar line generated by the essential matrix. Assuming independence, the overall likelihood of a motion hypothesis is proportional to the product of the likelihoods for all points:

$$P(E_i) \propto \Pi_{\mathbf{s}} P(E_i|\mathbf{s}) \tag{6}$$

After a dense correspondence probability distribution has been computed for all points, the method (Domke and Aloimonos, 2006) computes a probability distribution over motion hypotheses represented by the epipolar constraint. Finally, given the top ranked motion hypotheses, a Nelder-Mead simplex method (Lagarias et al., 1998) is used to refine the motion estimate.

However, since the current method does not allow motion scale recovery, translation T_s component does not contain image scale information. This type of information, is calculated by an alternative absolute orientation method like the Procrustes method.

2.2 Procrustes Analysis and Scale Factor Recovery

The Procrustes method allows to recover rigid body motion between frames, through the use of 3D point matches. We assume a set of 3D features (computed by triangulation of SIFT features) in instant t_{Tk} be described by $\{X_i'\}_{Tk}$, move to a new position and orientation in t_{Tk+1}, described by $\{Y_i'\}_{Tk+1}$. This transformation can be represented as:

$$Y_i' = RX_i' + T \tag{7}$$

where Y_i' points, are 3D feature points in I_{Tk+1}.

These points were detected using SIFT descriptors between I^L_{Tk} and I^L_{Tk+1}, that were triangulated to their stereo corresponding matches in I^R_{Tk+1}.

These two sets of points are the ones that are used by Procrustes method to estimate motion scale.

In order to estimate motion [R,T], a cost function that measures the sum of squared distances between corresponding points is used.

$$c^2 = \sum_{i}^{n} \left\| Y_i' - (RX_i' + T) \right\|^2 \quad (8)$$

Performing minimization of equation (8), gives estimates of $[R,T]$. Although conceptually simple, some aspects regarding the practical implementation of the Procrustes method must be taken into consideration. Namely, since this method is very sensible to data noise, obtained results tend to vary in the presence of outliers. To overcome this difficulty, RANSAC (Fischler and Bolles, 1981) is used to discard possible outliers within the set of matching points.

For a correct motion scale estimation, it is necessary to have a proper spatial feature distribution through out the image. For instance, if the Procrustes method uses all obtained image feature points without having their image spatial distribution into consideration, the obtained motion estimation [R,T] between two consecutive images could turn out biased.

Given these facts, to avoid having biased samples in the RANSAC phase of the algorithm, a bucketing technique (Zhang et al., 1995) is implemented to assure a unbiased image feature distribution sample. After, completing all this steps, only valid points are used in Procrustes method application. We then use an Extended Kalman filter to help robust camera linear and angular velocity estimates, and also to estimate vehicle pose between different time frames.

3 RESULTS

To illustrate the performance of our 6D Visual Odometry method, we compared our system performance against LIBVISO (Kitt et al., 2010), which is a standard library for computing 6 DOF motion. We also compared our performance against Inertial Measurement Unit (RTK-GPS information) using part of one of Kitt et al(Kitt et al., 2010) Karlsruhe dataset sequences.

In figure 3 one can observe angular velocity estimation from both IMU and LIBVISO, together with 6dp-RAW and EKF filtered measurements. All vision approaches obtained results are consistent with the Inertial Measurement Unit, but the 6dp-EKF displays a better performance in what respects the angular velocities. These results are stated in table (1), where root

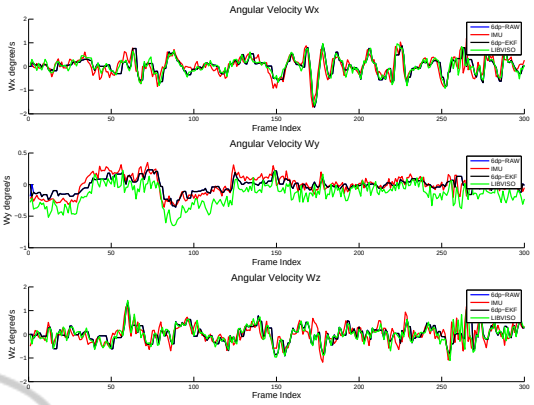

Figure 3: Angular Velocity Estimation Results.

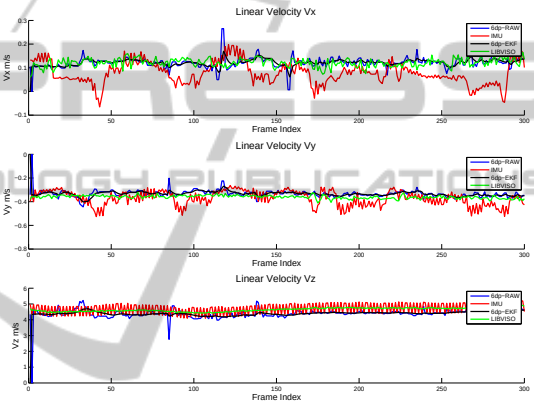

Figure 4: Linear Velocity Estimation Results.

mean square error between IMU and LIBVISO, 6DP-EKF estimation are displayed. Both methods display considerable low standard deviation results, but with 6DP-EKF displaying 50% less than LIBVISO for the angular velocities estimation.

Although not as good as the angular velocities, the 6dp-EKF method also displays a stable performance in obtaining linear velocity estimates using the sparse feature approach based on SIFT features combined with Procrustes Absolute Orientation method, as displayed in figure 4.

4 CONCLUSIONS AND FUTURE WORK

In this paper, we developed a novel method for conducting 6D visual odometry based on the use of dense Probabilistic Egomotion estimation approach. We also complemented this method with a sparse feature approach for estimating image depth. We tested the proposed algorithm against a state-of-the-art 6D vi-

Table 1: Standard Mean Squared Error between IMU and Visual Odometry (LIBVISO and 6dp-EKF).

	V_x	V_y	V_z	Ω_x	Ω_y	Ω_z
LIBVISO	0.0674	0.7353	0.3186	0.0127	0.0059	0.0117
6DP-EKF	0.0884	0.0748	0.7789	0.0049	0.0021	0.0056

sual Odometry Library such as LIBVISO.

The presented results demonstrate that 6DP performs accurately when compared to other techniques for 6-DOF visual Odometry estimation, yielding robust motion estimation results, mainly in the angular velocities estimation results.

In future work, we want to extend our dense probabilistic method to developed a standalone approach for ego-motion estimation that can cope with motion scale estimation, by using other type of multiple view geometry parametrization.

ACKNOWLEDGEMENTS

This work is financed by the ERDF â European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT Fundacao para a Ciencia e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP - 01-0124-FEDER-022701 and under grant SFRH / BD / 47468 / 2008 .

REFERENCES

Alcantarilla, P., Bergasa, L., and Dellaert, F. (2010). Visual odometry priors for robust EKF-SLAM. In *IEEE International Conference on Robotics and Automation,ICRA 2010*, pages 3501–3506. IEEE.

Civera, J., Grasa, O., Davison, A., and Montiel, J. (2010). 1-Point RANSAC for EKF filtering. Application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, 27(5):609–631.

Comport, A., Malis, E., and Rives, P. (2007). Accurate Quadri-focal Tracking for Robust 3D Visual Odometry. In *IEEE International Conference on Robotics and Automation, ICRA'07*, Rome, Italy.

Domke, J. and Aloimonos, Y. (2006). A Probabilistic Notion of Correspondence and the Epipolar Constraint. In *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pages 41–48. IEEE.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Huang, J., Zhu, T., Pan, X., Qin, L., Peng, X., Xiong, C., and Fang, J. (2011). A high-efficiency digital image correlation method based on a fast recursive scheme. *Measurement Science and Technology*, 21(3).

Kitt, B., Geiger, A., and Lategahn, H. (2010). Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. In *IEEE Intelligent Vehicles Symposium (IV), 2010*, pages 486–492. IEEE.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM J. on Optimization*, 9(1):112–147.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Maimone, M. and Matthies, L. (2005). Visual Odometry on the Mars Exploration Rovers. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 903–910. Ieee.

Moreno, F., Blanco, J., and González, J. (2007). An efficient closed-form solution to probabilistic 6D visual odometry for a stereo camera. In *Proceedings of the 9th international conference on Advanced concepts for intelligent vision systems*, pages 932–942. Springer-Verlag.

Ni, K., Dellaert, F., and Kaess, M. (2009). Flow separation for fast and robust stereo odometry. In *IEEE International Conference on Robotics and Automation ICRA 2009*, volume 1, pages 3539–3544.

Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26:756–777.

Nistér, D., Naroditsky, O., and Bergen, J. (2006). Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1):3–20.

Scaramuzza, D. and Fraundorfer, F. (2011). Visual odometry [tutorial]. *Robotics Automation Magazine, IEEE*, 18(4):80 –92.

Scaramuzza, D., Fraundorfer, F., and Siegwart, R. (2009). Real-time monocular visual odometry for on-road vehicles with 1-point ransac. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 4293 –4299.

Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q.-T. (1995). A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Special Volume on Computer Vision*, 78(2):87 – 119.