Proceedings of the 2007 IEEE/RSJ International
Conference on Intelligent Robots and Systems
San Diego, CA, USA, Oct 29 - Nov 2, 2007



# Vision-based Loop Closing for Delayed State Robot Mapping

Viorela Ila, Juan Andrade-Cetto, Rafael Valencia and Alberto Sanfeliu

*Abstract*— **This paper shows results on outdoor vision-based loop closing for Simultaneous Localization and Mapping. Our experiments show that for loops of over 50m, the pose estimates maintained with a Delayed-State Extended Information Filter are consistent enough to guarantee assertion of vision-based pose constraints for loop closure, provided no necessary information links are added to the estimator. The technique computes relative pose constraints via a robust least squares minimization of 3D point correspondences, which are in turn obtained from the matching of SIFT features over candidate image pairs. We propose a loop closure test that checks both for closeness of means and for highly informative updates at the same time.**

## I. Introduction

Closing large loops during Simultaneous Localization and Mapping (SLAM) is quite challenging. But, when accomplished, it reduces the accumulated estimation error enormously. A straight forward solution to the loop closing problem is to rely on the pose estimates from the filter of choice (be it a Kalman filter, an Information filter, or a particle filter) and perform data association tests as much and as often as possible. By testing for data association based on the likelihood of estimates, one can avoid some of the problems associated with appearance-based SLAM, such as aliasing for homogeneous or repetitive scenes. But, closing all loops consistent with the likelihood of estimates might add information links that are either unnecessary or unreliable.

*Unnecessary* links are those that contribute with little information to reduce the estimation error. That is the case with repetitive measurements to the same landmarks when little or no motion occurs, or when pose constraints from small steps are used in the case of a delayed state representation. *Unreliable* links on the other hand, are those that are not consistent with the sensor uncertainties for which the system was trained. This happens for example when distance estimates come from computer vision sensors. Distance errors being inverse to disparity in images are larger for measurements to far away landmarks, and their associated sensor covariances computed from linearized models tend to be optimistic.

Thus it is important to close loops sparsely on the one hand, and reliably on the other. In [1] the authors suggest that testing for loop closure in SLAM should be performed independently of the vehicle pose estimates. This is certainly a nice thing to do once the filter has become inconsistent. In this work, we adopt instead the straightforward approach of relying on the estimator for the generation of pose constraint hypotheses, but paying special attention no to let the system become overconfident too soon.

Since adding information links for all possible matches produces overconfident estimates that in the long term lead to filter inconsistency, we propose instead a two step loop closure test. First, we check whether two poses are candidates for loop closure with respect to their mean estimates. This is achieved by testing for the Mahalanobis distance in the same way data association gating is commonly performed during a SLAM update. But instead of adding all these information links, we limit the candidates to a second test that checks for large values on the second term of the Bhattacharyya distance. The purpose of this part of the test is to allow loop closing only on highly informative situations. That is, when the proposed matching covariances are sufficiently different, and a large amount of information is expected to enter the filter.

The reminder of this paper is as follows. In section II we present our strategy for computing pose constraints from two vantage points using computer vision. Section III is devoted to explain our chosen SLAM representation, and the way in which our vision-based pose constraints are used to update the map. Section IV described the proposed loop closure test. Some experimental results in an outdoor scenario are shown in Section V, and concluding remarks are added in Section VI.

## II. Relative Pose Constraints

The technique iterates as follows: SIFT image features are extracted and matched from candidate stereo image pairs. Their image point correspondences are then triangulated to obtain a set of 3D feature matches, which are in turn used to compute a least squares best fit pose transformation. Robust feature outlier rejection is obtained via RANSAC during the computation of the best camera pose constraint. These camera pose constraints are used as relative pose measurements in a delayed-state information-form SLAM. A substantial computational complexity advantage of the delayed-state information-form SLAM is that predictions and updates take constant time prior to loop closure given its exact sparseness [2]. Thanks to the features used, the proposed technique is robust enough not only to relate consecutive image pairs during robot motion, but also, to assert loop closure hypotheses.

### A. Feature Extraction

Simple correlation-based features, such as Harris corners [3] or Shi and Tomasi features [4], are of common use in

The authors are with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC. Llorens Artigas 4-6, Barcelona, 08028 Spain. [vila,cetto,rvalenci,sanfeliu]@iri.upc.edu.

1-4244-0912-8/07/$25.00 ©2007 IEEE.

vision-based SFM and SLAM; from the early uses of Harris himself to the popular work of Davison [5]. This kind of features can be robustly tracked when camera displacement is small and are tailored to real-time applications. However, given their sensitivity to scale, their matching is prone to fail under larger camera motions; less to say for loop-closing hypotheses testing. Given their scale and local affine invariance properties, we opt to use SIFTs instead [6], [7], as they constitute a better option for matching visual features from varying poses. To deal with scale and affine distortions in SIFTs, keypoint patches are selected from difference-of-Gaussian images at various scales, for which the dominant gradient orientation and scale are stored.

In our system, image pairs are acquired from a calibrated stereo rig[1]. Features are extracted and matched with previous image pairs. The surviving features are then stereo triangulated enforcing epipolar and disparity constraints. The epipolar constraint is enforced by allowing feature matches only within $\pm 1$ pixel rows on rectified images. The disparity constraint is set to allow matches within a $1-10$ meter range, where camera resolution is best. The result is a set of two clouds of matching 3D points $\mathbf{p}_t$ from the current pose, and $\mathbf{p}_i$ from a previous pose, $0 < i < t$, both referenced to the coordinate frame of the left camera.

### B. Pose Estimation

The homogeneous transformation relating the two aforementioned clouds of points can be computed by solving a set of equations of the form

$$\mathbf{p}_t = \mathbf{R}\mathbf{p}_i + \mathbf{t}. \tag{1}$$

A solution for the rotation matrix $\mathbf{R}$ is computed by minimizing the sum of the squared errors between the rotated directional vectors[2] of feature matches for the two robot poses. The solution to this minimization problem gives an estimate of the orientation of one cloud of points with respect to the other, and can be expressed in quaternion form as

$$\frac{\partial}{\partial \mathbf{R}}\left(\mathbf{q}^\top \mathbf{A}\mathbf{q}\right) = 0, \tag{2}$$

where $\mathbf{A}$ is given by

$$\mathbf{A} = \sum_{k=1}^{N} \mathbf{B}_k \mathbf{B}_k^\top, \tag{3}$$

$$\mathbf{B}_k = \begin{bmatrix} 0 & -c_x^k & -c_y^k & -c_z^k \\ c_x^k & 0 & b_z^k & -b_y^k \\ c_y^k & -b_z^k & 0 & b_x^k \\ c_z^k & b_y^k & -b_x^k & 0 \end{bmatrix}, \tag{4}$$

and

$$\mathbf{b}^k = \mathbf{v}_t^k + \mathbf{v}_i^k, \quad \mathbf{c}^k = \mathbf{v}_t^k - \mathbf{v}_i^k \quad . \tag{5}$$

The quaternion $\mathbf{q}$ that minimizes the argument of the derivative operator in the differential Equation 2 is the smallest eigenvector of the matrix $\mathbf{A}$.

[1]Point Gray's Bumblebee firewire stereo $640 \times 480$ camera, with a 6mm lens.

[2]A directional vector $\mathbf{v}$ can be computed as the unit norm direction along $\mathbf{p}$, and indicates the orientation of such point.

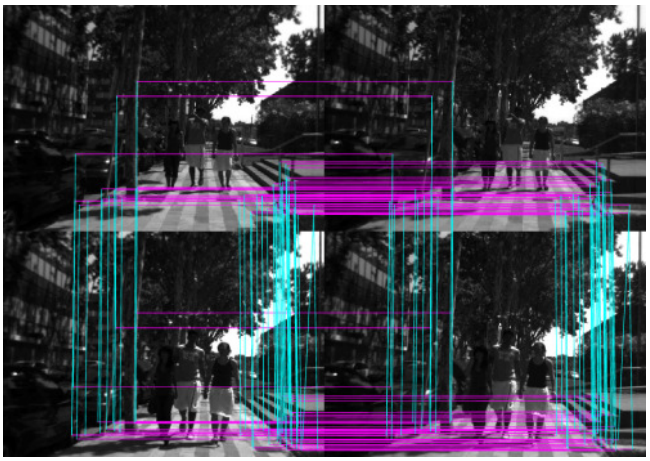

Fig. 1. SIFT correspondences in two consecutive stereo image pairs after outlier removal using RANSAC.

If we denote this smallest eigenvector by the 4-tuple $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$, it follows that the rotational angle $\theta$ associated with the rotational transform is given by

$$\theta = 2cos^{-1}(\alpha_1), \tag{6}$$

and the axis of rotation would be given by

$$\hat{\mathbf{a}} = \frac{(\alpha_2, \alpha_3, \alpha_4)^\top}{\sin(\theta/2)}. \tag{7}$$

Then, it can be shown that the elements of the rotation submatrix $\mathbf{R}$ are related to the orientation parameters $\hat{\mathbf{a}}$ and $\theta$ by

$$\mathbf{R} = \begin{bmatrix} a_x^2 + (1-a_x^2)c_\theta & a_x a_y c'_\theta - a_z s_\theta & a_x a_z c'_\theta + a_y s_\theta \\ a_x a_y c'_\theta + a_z s_\theta & a_y^2 + (1-a_y^2)c_\theta & a_y a_z c'_\theta - a_x s_\theta \\ a_x a_z c'_\theta - a_y s_\theta & a_y a_z c'_\theta + a_x s_\theta & a_z^2 + (1-a_z^2)c_\theta \end{bmatrix}, \tag{8}$$

where $s_\theta = \sin\theta$, $c_\theta = \cos\theta$, and $c'_\theta = 1 - \cos\theta$ [8].

Once the rotation matrix $\mathbf{R}$ is computed, we can use again the matched set of points to compute the translation vector $\mathbf{t}$

$$\mathbf{t} = \sum_{k=1}^{N} \mathbf{p}_i^k - \mathbf{R} \sum_{k=1}^{N} \mathbf{p}_t^k \quad . \tag{9}$$

It might be the case that SIFT matches occur on areas of the scene that experienced motion during the acquisition of the two image stereo pairs. For example, an interest point might appear at an acute angle of a tree leaf shadow, or on a person walking in front of the robot. The corresponding matched 3D points will not represent good fits to the camera motion model, and might introduce large bias to our least squares pose error minimization. To eliminate such *outliers*, we resort to the use of RANSAC [9]. The use of such a robust model fitting technique allows us to preserve the largest number of point matches that at the same time minimize the square sum of the residuals $\|\mathbf{R}\mathbf{p}_t + \mathbf{t} - \mathbf{p}_i\|$, as shown in Figure 1.

$$\mathbf{H} = \begin{bmatrix} \frac{\Delta x\cos\psi+\Delta y\sin\psi}{d} & \frac{-\Delta x\sin\psi+\Delta y\cos\psi}{d} & 0 & \frac{-\Delta x\cos\psi-\Delta y\sin\psi}{d} & \frac{\Delta x\sin\psi-\Delta y\cos\psi}{d} & d\sin\psi \\ \frac{\Delta x\sin\psi-\Delta y\cos\psi}{d} & \frac{\Delta x\cos\psi+\Delta y\sin\psi}{d} & 0 & \frac{-\Delta x\sin\psi+\Delta y\cos\psi}{d} & \frac{-\Delta x\cos\psi-\Delta y\sin\psi}{d} & -d\cos\psi \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \quad (20)$$

## III. Visually Augmented Odometry in Information Form

### A. Exactly Sparse Delayed-State SLAM

Compared to the Extended Kalman Filter (EKF) SLAM which has quadratic time complexity in the number of states, the delayed-state information-form SLAM has been shown to produce exactly sparse information matrices [2]. If consecutive robot poses are added to the state, the result is a tri-block diagonal information matrix, linking consecutive measurements. This situation allows constant predictions and updates, with a considerable advantage in terms of computational cost and making it suitable for relative large environments. This three-block diagonal structure is only modified sparsely during loop closure. Thus, to keep the computational complexity low, we would like to only close loops when these are really needed. Our loop closure test is meant to do exactly that.

The delayed-state information-form SLAM representation consists on estimating a state vector $\mathbf{x}$ with the history of poses parameterized as an inverse normal distribution.

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x};\mu,\Sigma) = \mathcal{N}^{-1}(\mathbf{x};\eta,\Lambda), \quad (10)$$

where

$$\Lambda = \Sigma^{-1} \qquad \text{and} \qquad \eta = \Lambda\mu \quad (11)$$

In a delayed state representation, the map state is simply the history of pose mean estimates

$$\mu = \begin{bmatrix} \mu_t \\ \vdots \\ \mu_1 \end{bmatrix}.$$

As in most SLAM formulations, white noise $\mathbf{w}_t$ with covariance $\mathbf{Q}$ is added to the vehicle motion prediction model, and its linearized version is used in the computation of covariance prediction (information prediction in our case).

$$\begin{aligned} \mathbf{x}_{t+1} &= f(\mathbf{x}_t,\mathbf{u}_t) + \mathbf{w}_t \\ &\approx f(\mu_t,\mathbf{u}_t) + \mathbf{F}(\mathbf{x}_t - \mu_t) + \mathbf{w}_t \end{aligned}. \quad (12)$$

Given the absolute dead-reckoning readings of position $x^d$, $y^d$, and orientation $\theta$, the delayed state can be written as

$$\mathbf{x} = [x_t, y_t, \theta_t, \ldots x_1, y_1, \theta_1]^\top. \quad (13)$$

The vehicle motion model can be express in terms of the relative travelled distance and relative orientation change

$$x_{t+1} = x_t + d\cos(\theta_t + \psi) \quad (14)$$

$$y_{t+1} = y_t + d\sin(\theta_t + \psi) \quad (15)$$

$$\theta_{t+1} = \theta_t + \Delta\theta. \quad (16)$$

At two consecutive time steps $t$ and $t+1$, the relative travelled distance and relative orientation change are

$$d = \sqrt{(x^d_{t+1} - x^d_t)^2 + (y^d_{t+1} - y^d_t)^2} \quad (17)$$

$$\psi = \tan^{-1}\left(\frac{y^d_{t+1} - y^d_t}{x^d_{t+1} - x^d_t}\right) - \theta_t \quad (18)$$

$$\Delta\theta = \theta_{t+1} - \theta_t. \quad (19)$$

Considering the linearization of the nonlinear motion prediction model from Equation 12 we can calculate the Jacobian $\mathbf{F}$ as

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & -d\sin(\theta_t + \psi) \\ 0 & 1 & d\cos(\theta_t + \psi) \\ 0 & 0 & 1 \end{bmatrix}. \quad (21)$$

The revision of the entire history of poses, as a result of adding the new information that links the current and predicted poses, can be computed in information form [2] with

$$\bar{\eta} = \begin{bmatrix} \mathbf{Q}^{-1}(\mathbf{f}(\mu_t,\mathbf{u}_t) - \mathbf{F}\mu_t) \\ \eta_t - \mathbf{F}^\top\mathbf{Q}^{-1}(\mathbf{f}(\mu_t,\mathbf{u}_t) - \mathbf{F}\mu_t) \\ \eta_{t-1:1} \end{bmatrix}, \quad (22)$$

and the associated information matrix is

$$\bar{\Lambda} = \begin{bmatrix} \mathbf{Q}^{-1} & -\mathbf{Q}^{-1}\mathbf{F} & \mathbf{0} \\ -\mathbf{F}^\top\mathbf{Q}^{-1} & \Lambda_{t,t} + \mathbf{F}^\top\mathbf{Q}^{-1}\mathbf{F} & \Lambda_{t,t-1} \\ \mathbf{0} & \Lambda_{t-1:1,t} & \Lambda_{t-1:1,t-1:1} \end{bmatrix}, \quad (23)$$

Augmenting the information vector in this form introduces shared information only between the new robot pose $\mathbf{x}_{t+1}$ and the previous one $\mathbf{x}_t$. Moreover, the shared information between $\mathbf{x}_{t+1}$ and the delayed-states ($t-1$ to 1) is always zero, resulting in a naturally sparse information matrix with a tridiagonal block structure.

Now, in the information-form measurement updates are additive and can be computed in constant-time. The pose constraints linking the predicted pose and any previous nearby poses, as measured by our vision system would have the form

$$\begin{aligned} \mathbf{z}_{t+1} &= \mathbf{h}(\mathbf{x}_{t+1,i}) + \mathbf{v}_t \\ &\approx \mathbf{h}(\bar{\mu}_{t+1,i}) + \mathbf{H}(\mathbf{x}_{t+1,i} - \bar{\mu}_{t+1,i}) + \mathbf{v}_{t+1} \end{aligned}, \quad (24)$$

with $\mathbf{v}_{t+1}$ the zero mean, white measurement noise with covariance $\mathbf{R}$ and the measurement Jacobian $\mathbf{H}$.

The nonlinear measurement model is

$$z_x = d\cos(\psi) \quad (25)$$

$$z_y = d\sin(\psi) \quad (26)$$

$$z_\theta = \theta_{t+1} - \theta_i. \quad (27)$$

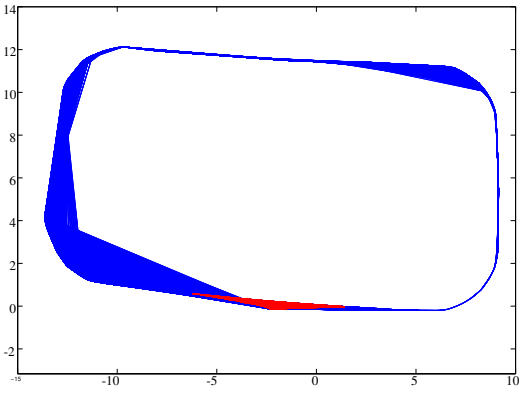

Fig. 2. Links coming from Mahalanobis (blue) and Mahalanobis-Bhattacharyya distance (red) tests.

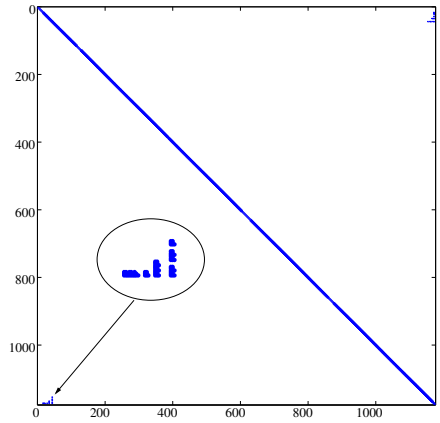

Fig. 3. Sparse information matrix in delayed state SLAM. Loop-closure event adds non-zero off-diagonal elements (see zoomed region).

where

$$d = \sqrt{(x_{t+1}-x_i)^2+(y_{t+1}-y_i)^2} \quad (28)$$

$$\psi = \tan^{-1}\left(\frac{y_{t+1}-y_i}{x_{t+1}-x_i}\right)-\theta_i \quad (29)$$

and the Jacobian $\mathbf{H}$ is computed as shown in Equation 20 with

$$\Delta x = x_{t+1}-x_i \quad (30)$$

$$\Delta y = y_{t+1}-y_i\,. \quad (31)$$

The update to the entries linking poses $t+1$ and $i$ in the information vector and information matrix become:

$$\eta_{t+1,i} = \bar{\eta}_{t+1,i}+\mathbf{H}^\top\mathbf{R}^{-1}(\mathbf{z}_{t+1}-\mathbf{h}(\bar{\mu}_{t+1,i})+\mathbf{H}\bar{\mu}_{t+1,i}) \quad (32)$$

$$\Lambda_{t+1,i,t+1,i} = \bar{\Lambda}_{t+1,i,t+1,i}+\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}\,. \quad (33)$$

$\mathbf{R}$ now represents a covariance term for the sensor uncertainty. It can be computed from the first order Taylor series expansion of the stereo correspondence geometry, computed from the Maximum Likelihood Estimate and the assumed pixel feature noise [10]. A more elaborate derivation might also include the uncertainty of the SIFT feature extractor. In this work however, we have chosen to let $\mathbf{R}$ be a fixed value, leaving its more elaborate derivation to future research. The reason is that the linearization of the stereo correspondence geometry with respect to the pixel feature noise, required in a the computation of $\mathbf{R}$, will only delay but not prevent the filter from becoming inconsistent.

The Jacobian $\mathbf{H}$ is always sparse [11] and as a consequence only the four blocks relating poses $t+1$ and $i$ in the information matrix will be updated

$$\mathbf{H} = [\mathbf{H}_{t+1},0,\ldots,0,\mathbf{H}_i,0,\ldots,0]\,. \quad (34)$$

For consecutive poses, this produces tridiagonal blocks in $\Lambda$, and for larger loop closures, the updates only modify the diagonal $t+1$ and $i$ terms and introduces sparse blocks at the locations $t+1,i$ and $i,t+1$.

## IV. Loop Closure Detection

Two phases can be distinguished during the loop-closing process. First, we need to detect the possibility of a loop closure event and then, we must certify the presence of such loop closure from visual data. The likelihood of pose estimates are valuable in detecting possible loop closures. A comparison of the current pose estimate with the history of poses can tell whether the robot is in the vicinity of a previously visited place, in terms of both the global position and the orientation. This is achieved by measuring the Mahalanobis distance from the prior estimate to all previously visited locations, i.e., for all $0<i<t$,

$$d_M^2 = (\mu_{t+1}-\mu_i)^\top\left(\frac{\Sigma_{t+1}+\Sigma_i}{2}\right)^{-1}(\mu_{t+1}-\mu_i) \quad (35)$$

An exact computation of $\Sigma_{t+1}$ and $\Sigma_i$ requires the inverse of $\bar{\Lambda}$, which can be computed in linear time using conjugate gradient techniques [2]. Motivated by [11], these covariances can be efficiently approximated in constant time from their Markov blankets. Note also that Eq. 35 does not take into account the cross correlation between poses in the Mahalanobis metric, but this can be done with no substantial extra effort. The only difference is that instead of computing individual Markov blankets for each pose, the combined Markov blanket is used.

The average covariance is used to accommodate for the varying levels of estimation uncertainty both on the pose prior being evaluated, and on the past pose being compared. In case of a normal distribution, the Mahalanobis distance follows the $\chi^2$-square distribution with n-1 degrees of freedom (where n is the number of variables; in our case, 2dof on a 95% confidence bound suffice).

Many nearby poses will satisfy this condition, as shown in Figure 2. At the start of a SLAM run, when covariances are small, only links connecting very close poses will satisfy the test. But, as error accumulates, pose covariances grow covering larger and larger areas of matching candidates. For long straight trajectories having corresponding visual

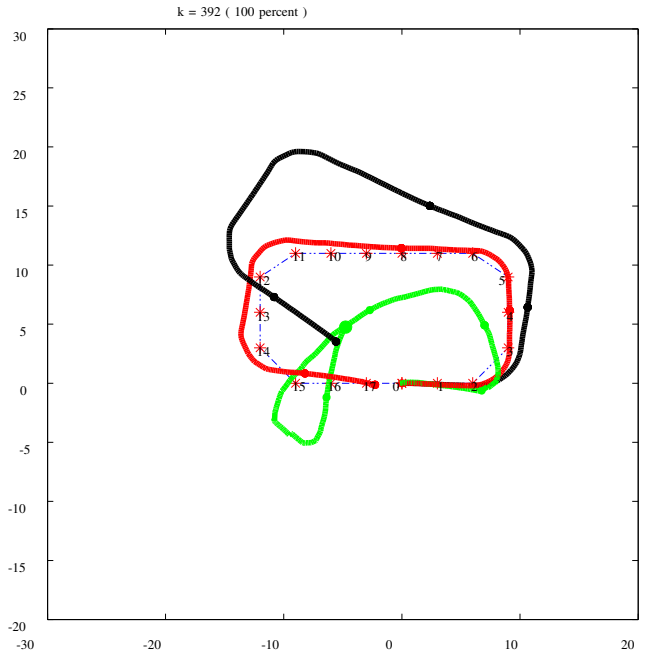


Fig. 4. Odometry-only (black), vision-only (green), and combined vehicle trajectory (red).

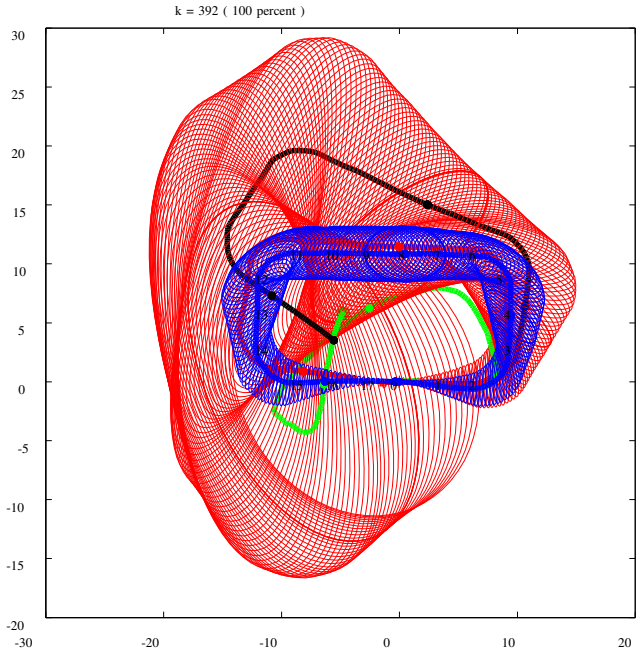


Fig. 5. The information added during the loop closure event enormously reduces the accumulated errors: prediction (red), update (blue).

features, information links could even be established. Some aid in reducing the effect of large pose constraints, because orientation variance is usually larger than position variance, and pose covariance ellipsoids usually align orthogonal with respect to the direction of motion.

Due to linearization effects, adding information links for all possible matches produces overconfident estimates that in the long run lead to filter inconsistency. Thus, our update procedure must pass a second test. The aim of this second test is to allow updating using only links with high informative load. In terms of covariances, this happens when a pose with a large covariance can be linked with a pose with a small uncertainty.

$$d_B = \frac{1}{2} ln \frac{\left| \frac{\Sigma_{t+1} + \Sigma_i}{2} \right|}{\sqrt{| \Sigma_{t+1} || \Sigma_i |}} \tag{36}$$

The above expression refers to the second term of the Bhattacharyya distance, and gives a measure of separability in terms of covariance difference [12]. This test is typically used to discern between to distinct classes with close means but varying covariances. We can see however that it also can be used to fuse two observations of the same event with varying covariance estimates. Given that, the value of $d_B$ increases as the two covariances $\Sigma_{t+1}$ and $\Sigma_i$ are more different. The Bhattacharyya covariance separability measure is symmetric, and we need to test whether the current pose covariance is larger than the $i$-th pose it is being compared with. This is done by analyzing the area of uncertainty of each estimate by comparing the determinants of $| \Sigma_{t+1} |$ and $| \Sigma_i |$. The reason is that we only want to update the overall estimate with information links to states that had smaller uncertainty than the current state. Figure 2 shows in red the remaining links after the second test.

In a second phase we still must certify the presence of a loop closure event to update the entire pose estimate and to reduce overall uncertainty. When an image correspondence can be established, the computed pose constraint is used in a one-step update of the information filter, as shown in Equations 33 and 32. A one-step update in information form changes the entire history of poses adding a linear number of non-zero off-diagonal elements in the information matrix as shown in the Figure 3. The sparsity can be controlled by reducing the confidence on image registration when testing for loop-closure event.

## V. Experiments

To test our strategy for vision-based augmented odometry we have performed a series of experiments on urban unstructured environments of small size. In one of our experiments we chose a cyclic path of around 300 sq meters with the purpose to test a loop-closure situation. A snapshot of these tests is shown in Figure 4. The robot was manually driven through a series of predefined points previously marked on the floor. The results of estimating the vehicle motion purely from accumulated raw odometry and purely from concatenating vision pose constraints are shown in the figure as black and green plots, respectively. The delayed-state information-based revised trajectory resulting from the fusion of the two is shown in red. No motion was accumulated for those cases when not enough SIFT points were obtained during the computation of vision-based pose constraints. The effect of noninformative vision-based poses at some iterations can be efficiently modelled in our approach only by computing

motion predictions from odometry without performing map updates.

Due to the fact that raw odometry is really poor especially when the vehicle turns and the vision-based pose constraints can fail in translation estimation but provides quite accurate rotation estimation, our SLAM gives more weight to the translation measurements provided by the odometry and to the rotation estimated using SIFT points. Whereas odometry error accumulates monotonically, vision-based pose constraints vary in accuracy from very accurate (about 3,4 cm) to as large as 1 meter in estimation error. Nontheless the fusion of both provides a consistently revised pose estimate.

At each step both distances (Mahalanobis and Bhattacharyya) are used to test the presence of loop closure events. When a loop-closing event is announced, visual matching is performed among the corresponding indexed images. If sufficient point matches are found, the homogeneous transformation relating the pose constraint is used to update the history of poses, and their associated information terms. Figure 5 shows in blue the correction of the entire history of poses after a link of about 50 meters is established. Also shown in red, the covariance estimates prior to the update as maintained by the filter in information form.

## VI. Conclusions and Future Work

This paper proposed an efficient approach to vision-based loop closing problem in delayed-state robot mapping based on analyzing the similarity and separability of pose estimates from the mean and covariance difference measures. The mapping process is based on an exactly sparse delayed-state filter that uses SIFT features to compute pose constraints. Another type of features that we seek to explore in the future are *Speed Up Robust Features* SURFs [13]. These features have similar response properties to SIFTs, replacing Gaussian convolutions with Haar convolutions, and a significant reduction in computational cost.

We can conclude saying that, concerning the accuracy of the pose estimation, our approach performs well when comparing the estimated trajectory with ground truth points, and considerably reduces the memory and execution time by using a sparse information matrix to link consecutive measurements each time (time and memory increase linearly compared to the quadratic cost of the traditional EKF).

On the other hand, concerning the loop closure process, the method introduced here is a straightforward approach consistent with the likelihood of estimates. Instead of searching for visual correspondences within the whole history of images, we restrict our search by applying conservative tests for loop-closure event detection in terms of mean closeness and covariances separability. The experimental results show that these tests avoid unnecessary and unreliable links, substantially reducing the search time. Moreover, the established links enormously reduce the accumulated errors from the history of poses as shown in Figure 5 by adding only high informative elements to the filter.

The experimental results presented here constitute a preliminary study on the use of 6D vision-based pose transforms for outdoor mapping. We are currently working in testing this vision-based technique for mapping of wider areas, in the order of 10000 sq. meters, where robust loop closing methods as the one presented in this paper are necessary to reduce the accumulated errors in the mapping process. Our focus now is on pursuing strategies for intelligent pruning of the number of states in the filter, so as to keep the problem tractable for large areas.

## Acknowledgements

This research is supported by the Spanish Ministry of Education and Science under a Juan de la Cierva Postdoctoral Fellowship to V. Ila in project DPI 2004-07358, and a Ramón y Cajal Postdoctoral Fellowship to J. Andrade-Cetto, by a scholarship from the Mexican Council of Science and Technology to R. Valencia, the NAVROB Project DPI 2004-5414, and the EU URUS project FP6-IST-045062.

## References

[1] P. Newman and K. Ho, "SLAM loop-closing with visually salient features," in *Proc. IEEE Int. Conf. Robot. Automat.*, Barcelona, Apr. 2005, pp. 635–642.

[2] R. Eustice, H. Singh, and J. Leonard, "Exactly sparse delayed-state filters for view-based SLAM," *IEEE Trans. Robot.*, vol. 22, no. 6, pp. 1100–1114, Dec. 2006.

[3] C. G. Harris and M. Stephens, "A combined corner edge detector," in *Proc. Alvey Vision Conf.*, Manchester, Aug. 1988, pp. 189–192.

[4] J. Shi and C. Tomasi, "Good features to track," in *Proc. 9th IEEE Conf. Comput. Vision Pattern Recog.*, Seattle, Jun. 1994, pp. 593–600.

[5] A. J. Davison, I. Reid, N. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Machine Intell.*, 2007, to appear.

[6] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 10, pp. 1615–1630, 2005.

[8] B. K. P. Horn, H. M. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *J. Opt. Soc. Am. A*, vol. 5, no. 7, pp. 1127–1135, 1988.

[9] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, pp. 381–385, 1981.

[10] O. Faugeras, *Three-Dimensional Computer Vision. A Geometric Viewpoint*. Cambridge: The MIT Press, 1993.

[11] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous localization and mapping with sparse extended information filters," *Int. J. Robot. Res.*, vol. 23, no. 7-8, pp. 693–716, Jul. 2004.

[12] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego: Academic Press, 1990.

[13] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. 9th European Conf. Comput. Vision*, ser. Lect. Notes Comput. Sci., vol. 3951. Graz: Springer-Verlag, 2006, pp. 404–417.