Interacting Multiple Model Monocular SLAM

Javier Civera, Andrew J. Davison and J. M. M. Montiel

***Abstract*— Recent work has demonstrated the benefits of adopting a fully probabilistic SLAM approach in sequential motion and structure estimation from an image sequence. Unlike standard Structure from Motion (SFM) methods, this 'monocular SLAM' approach is able to achieve drift-free estimation with high frame-rate real-time operation, particularly benefitting from highly efficient active feature search, map management and mismatch rejection.**

A consistent thread in this research on real-time monocular SLAM has been to reduce the assumptions required. In this paper we move towards the logical conclusion of this direction by implementing a fully Bayesian Interacting Multiple Models (IMM) framework which can switch automatically between parameter sets in a dimensionless formulation of monocular SLAM. Remarkably, our approach of full sequential probability propagation means that there is no need for penalty terms to achieve the Occam property of favouring simpler models — this arises automatically. We successfully tackle the known stiffness in on-the-fly monocular SLAM start up without known patterns in the scene. The search regions for matches are also reduced in size with respect to single model EKF increasing the rejection of spurious matches. We demonstrate our method with results on a complex real image sequence with varied motion.

I. INTRODUCTION

A. Real-time sequential SFM estimation from sequences

Camera motion and scene structure estimation from an image sequence of a previously unknown scene has most often been performed as an off-line optimisation procedure (e.g. [9]), but with increasing computing power there have been several successful recent real-time algorithms. Real-time operation requires sequential processing with bounded computational requirements per frame, and there have been two key paradigms for achieving this. Firstly, algorithms which we can generically describe as *visual odometry* approaches sequentially determine motion and structure by concatenating estimates from sliding windows of two or more time-steps (e.g. [15], [17]) to produce arbitrarily long trajectories with constant-time processing cost. While this approach, which 'forgets' about the past, leads to motion estimates which drift over time, the rates of drift can be made extremely low if a great number of features are matched from frame to frame.

The second main approach is to use probabilistic filtering to recursively estimate a full probability density over the current camera pose and the positions of features — adopting the core Simulaneous Localisation and Mapping (SLAM) approach of the mobile robotics literature. If the number of features is restricted, and camera motion limited to a certain volume, this leads to systems which also have constant-time computation and can run in real-time to build consistent maps and estimate motions without drift. Successful monocular SLAM examples have used the Extended Kalman Filter (e.g. [7]) or Rao-Blackwellized particle filtering ([8]).

In this paper we follow the probabilistic filtering approach, which is preferable in the very common scenario of loopy, repeated motion within a restricted area, and specifically use a full-covariance EKF to estimate the locations of the camera and features. Maintaining an always-up-to-date full PDF over motion and structure estimates has several attractive advantages. In particular, it allows prediction of measurements for highly efficient active image search and to confirm match hypotheses (data association), and also intelligent incremental map management [6].

Our goal in this paper is to provide a framework within which the approach of fully sequential probability propagation can be applied to *any image sequence*. This has so far not been possible because sequential filtering algorithms depend on assumptions and parameters which determine their behaviour. The system of Davison in [7] assumed camera motion of certain dynamics (in terms of expected linear and angular accelerations), a scene with a maximum feature depth of around 5m and some known scene information in the form of an initialisation target.

There has been significant recent work on removing the restrictions of Davison's original EKF algorithm. One important research direction was to permit the probabilistic use of low-parallax features, either recently initialised or at extreme scene depths. Davison's feature initialisation scheme using an auxiliary particle filter was improved on by Solà *et al.* [19] with a mixture of Gaussians method, and then by Eade and Drummond [8] and Montiel, Civera and Davison [14], [3], [4] with a new inverse depth parameterisation which can seamlessly cope with features at any depth, and is able to work without any known initial pattern in the scene. Following this thread of work is the approach of Civera *et al.* [2] who have formulated a completely dimensionless monocular SLAM algorithm. Using an inverse depth parameterisation, they removed metric and time scales from the SLAM state vector and tuning parameters from the filter to formulate the whole problem in terms of dimensionless values interpretable as quantities in image space. When such monocular SLAM algorithms are applied to real image sequences, it is worth noting the valuable role that Joint Compatibility testing [16] can play in rejecting spurious matches that might otherwise ruin the whole estimation (as shown by Clemente *et al.* in [5]).

Javier Civera and J. M. M. Montiel are with Departamento de Informática e Ingeniería de Sistemas, University of Zaragoza, Spain. {jcivera, josemari}@unizar.es

Andrew J. Davison is with the Department of Computing, Imperial College, London, UK. ajd@doc.ic.ac.uk

B. IMM monocular SLAM

Image sequence processing relies on camera motion models to robustly identify point matches. Off-line methods rely on geometrical models relating two or three images to compute matches. In [20] it is shown how different models should to be used at different parts of a general sequence to avoid degenerate geometries. This geometrical model selection has been extended to segment different motion models between image pairs or triplets [18], [11], [21].

In contrast to these two or three-view geometrical models, the probabilistic motion models used in SLAM are well suited to modelling long sequences of close images instead of discrete sets of images. However a single probabilistic model can similarly only deal with sequences which follow the prescribed model or processing will fail. In this work we extend the monocular SLAM method to deal with more than one probabilistic motion model, expanding the range of sequences compatible with the priors represented by a set of tuning parameters. We use a sequential Bayesian approach to model selection.

Thanks to Bayesian probability propagation, monocular SLAM with a general translating camera can deal with low parallax motions — such as rotations — provided that the camera re-observes map features whose locations are well-estimated as a result of parallax observed previously in the sequence, and so model switching is not a must in some cases where it would be in the off-line approaches. However, when monocular SLAM is initialised on-the-fly without a known scene pattern, model selection is an issue. If the camera initially undergoes a low parallax motion, no reliable estimation is possible. Any measurement noise may be considered parallax by the filter producing inconsistent depth estimates. We tackle this problem with model selection.

Multiple model methods are well known in maneuvering target tracking. An excellent and recent survey of this can be found in [13]. In our paper, we adapt to the SLAM problem the most widespread of those methods, Interacting Multiple Models (IMM), initially proposed by Blom in [1]. The IMM estimator is a suboptimal hybrid filter — that is, it estimates the continuous values of a proccess, and the discrete probabilities of a set of models — whose main features are: 1) It assumes that the system can jump between the members of a set of models, which is the case of our monocular SLAM estimation, and 2) It offers the best compromise between complexity and performance.

Thanks to the use of multiple models, the range of images that can be processed with a single system tuning is enlarged. We work with a bank of 7 models: one model of a stationary camera, three models of pure rotation motion (constant angular velocity) with different angular acceleration covariances, and three general translation + rotation models (constant velocity, constant angular velocity) with different angular and linear acceleration covariances. Via the Bayesian model selection of IMM, the system prefers simpler (less general) models where they fit the data. As a result, the search regions for the predicted image features are smaller than with a single model EKF. These reduced search regions increase mismatch rejection and reduce the processing cost of image search. Additionally, the computed probabilities per model allow the segmentation of a sequence into different models.

Section II discusses and formulates sequential Bayesian model selection. The Interacting Multiple Model approach to Bayesian model selection is detailed in III. Some details about the use of IMM in the SLAM problem are given in Section IV. Section V verifies the method using real imagery and shows how it deals with sequence bootstrap. Finally Section VI summarises the paper's conclusions.

II. BAYESIAN MODEL SELECTION FOR SEQUENCES

In standard single-model monocular SLAM algorithms, Bayes' rule combines at every step past estimation information with current image data. Given the background information I and the image data at current step D, the posterior probability density function for the set of parameters θ defining our model M is updated via Bayes' formula:

$$p(\theta|DMI) = p(\theta|MI)\frac{p(D|\theta MI)}{p(D|MI)} \ . \tag{1}$$

In this paper we consider cases where a single model M is not sufficient to cover all of the sequences we would like to track. Taking full advantage of the fully probabilistic estimation that our SLAM approach is performing, we formulate our multiple model problem in a Bayesian framework.

Consider, as Jaynes does in Chapter 20 of his book [10], a discrete set of models $\mathcal{M} = \{M^1, \ldots, M^r\}$ — rather than a single one — which might feasibly describe the assumptions of a sequential SFM process. We start by assigning initial scalar probabilities $P(M^1|I), \ldots, P(M^r|I)$ which represent prior belief about the different models based on background information I, and which are normalised to add up to one. If no prior information exists, these probabilities may well be assigned initially equal.

At each new image, where we acquire image measurements data D, we update the probability of each model according to Bayes' rule:

$$P(M^j|DI) = P(M^j|I)\frac{P(D|M^jI)}{P(D|I)} \tag{2}$$

In this expression, the first term is the probability of the model being correct given only the prior information. In the fraction, the numerator is the likelihood of obtaining the data given that the model is correct. The denominator is the normalizing constant, computation of which can be avoided when the posterior probabilities of a mutually-exclusive set of models are all computed, or alternatively cancels out when the ratio of posterior probabilities of different models is calculated.

So, what is the likelihood $P(D|M^jI)$ of the data given a model in a monocular SLAM system? It is simply the joint likelihood of all of the feature measurements in an image:

$$P\left(D|MI\right) = \frac{1}{\sqrt{2\pi\left|\mathtt{S}\right|}}\exp\left(-\frac{1}{2}\nu^\top \mathtt{S}^{-1}\nu\right) , \tag{3}$$

where

$$\nu = \mathbf{z} - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) \ , \quad (4)$$

$$\mathtt{S} = \mathtt{H}\left(\mathtt{F}\mathtt{P}_{k|k-1}\mathtt{F}^{\top} + \mathtt{G}\mathtt{Q}\mathtt{G}^{\top}\right)\mathtt{H}^{\top} + \mathtt{R} \ . \quad (5)$$

$\mathtt{F} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}$, $\mathtt{B} = \frac{\partial \mathbf{f}}{\partial \mathbf{w}}$, $\mathtt{H} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}$ and $\mathtt{G} = \frac{\partial \mathbf{h}}{\partial \mathbf{u}}$ are the Jacobians for the EKF.

We should note, as Jaynes explains with great clarity, that in this correctly formulated Bayesian approach to model selection there is no need for ad-hoc terms like Minimum Description Length which penalise 'complex' models and favour simple ones. The 'Occam principle' of selecting the simplest model which is able to capture the detail of the data and avoiding overfitting is taken care of automatically by correctly normalising our comparison of different models. The big difference between our approach and the common two-view model selection methods (e.g. [11], [21], [18]) which require penalty terms is that our concept of a model is probabilistic at its core, not just geometric (like homography, affine, ...). For our use in sequential probabilistic tracking, a model must actually define a probability distribution during a transition. This is what makes it possible to calculate proper likelihoods for the models themselves, independent of parameters.

The formulation above allows us to obtain posterior probabilities for our models in one frame, but we are interested in propagating these probabilities through a sequence. This is achieved by defining a vector of model probabilities — a 'state vector' for models or set of mixing weights:

$$\mu_{\mathbf{k}|\mathbf{k}} = \left(\mu^{\mathbf{1}}_{\mathbf{k}|\mathbf{k}} \cdots \mu^{\mathbf{r}}_{\mathbf{k}|\mathbf{k}}\right)^{\top} \ . \quad (6)$$

We fill $\mu_{\mathbf{k}|\mathbf{k}}$ with the prior model probabilities $P(M^1|I), \ldots, P(M^r|I)$ before processing the first image, and after processing use the values of $\mu_{\mathbf{k}|\mathbf{k}}$ as the priors for each model in Equation 2 and then replace these values with the posterior values calculated.

A final step is needed in between processing images, which is to apply a mixing operator to account for possible transitions between models. With a homogeneous Markov assumption that the probability of transition from one model to any other is constant at any inter-frame interval, this is achieved by:

$$\mu_{\mathbf{k}|\mathbf{k-1}} = \pi \mu_{\mathbf{k-1}|\mathbf{k-1}} \ , \quad (7)$$

where π is a square matrix of transition probabilities where each row must be normalised. In the typical case that the dominant tendancy is sustained periods of motion with one model, this matrix will have large terms on the diagonal. If the models are ordered with some sense of proximity, the matrix will tend to have large values close to the diagonal and small ones far away.

The sequential process of calculating model probabilities therefore evolves as a loop of mixing and update steps and at motion transitions in the sequence evidence will accrue over several frames.

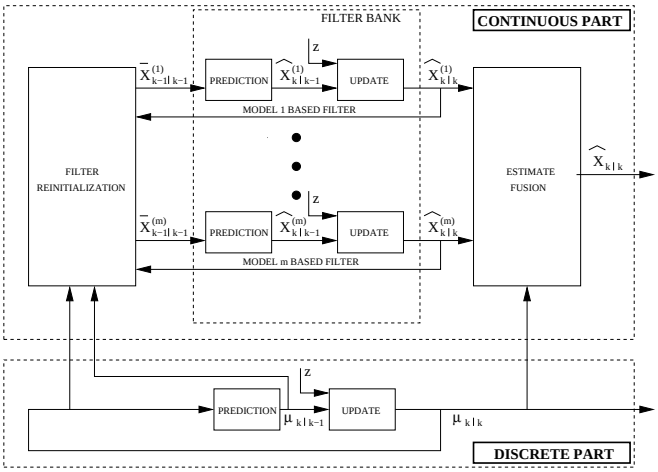

Fig. 1. Interacting Multiple Model algorithm scheme

1. Filter reinitialization (for $i = 1, 2, ..., r$):
`Predicted model probability:`
$\mu^i_{k|k-1} = P\{M^i_k|z^{k-1}\} = \sum_j \pi_{ji}\mu^j_{k-1}$
`Mixing weight:`
$\mu^{j|i}_{k-1} = P\{M^j_{k-1}|M^i_k, z^{k-1}\} = \pi_{ji}\mu^j_{k-1}/\mu^i_{k|k-1}$
`Mixing estimate:`
$\bar{\mathbf{x}}^i_{k-1|k-1} = E[\mathbf{x}_{k-1}|m^i_k, z^{k-1}] = \sum_j \mathbf{x}^j_{k-1|k-1}\mu^{j|i}_{k-1}$
`Mixing covariance:`
$\bar{P}^i_{k-1|k-1} = \sum_j (P^j_{k-1|k-1} +$
$+(\bar{\mathbf{x}}^i_{k-1|k-1} - \hat{\mathbf{x}}^j_{k-1|k-1})(\bar{\mathbf{x}}^i_{k-1|k-1} - \hat{\mathbf{x}}^j_{k-1|k-1})^{\top})\mu^{j|i}_{k-1}$

2. EKF bank filtering (for $i = 1, 2, ..., r$):
`Prediction:` $\hat{\mathbf{x}}^i_{k|k-1}, P^i_{k|k-1}, \mathbf{h}(\mathbf{x}^i_{k|k-1}), S^i_k$
`Measurement:` z_k
`Update:` $\hat{\mathbf{x}}^i_{k|k}, P^i_{k|k}$

3. Model probability update (for $i = 1, 2, ..., r$):
`Model likelihood:` $L^i_k = \mathcal{N}(\nu^i_k; 0, S^i_k)$
`Model probability:` $\mu^i_k = \frac{\mu^i_{k|k-1}L^i_k}{\sum_j \mu^j_{k|k-1}L^j_k}$

4. Estimate fusion
`Overall state:`
$\hat{\mathbf{x}}_{k|k} = \sum_i \hat{\mathbf{x}}^i_{k|k}\mu^i_k$
`Overall covariance:`
$P_{k|k} = \sum_i \left(P^i_{k|k} + (\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}^i_{k|k})(\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}^i_{k|k})^{\top}\right)\mu^i_k$

Fig. 2. Interacting Multiple Model algorithm

III. INTERACTING MULTIPLE MODEL

IMM is presented in the tracking literature as a hybrid estimation scheme, well suited to estimating the continuous state of a system that can switch between several behaviour modes. This hybrid system is then composed of a continuous part (the state) and a discrete part (the behaviour modes). The continuous part of such a system is defined by its state and measurement equations:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), M(t), \mathbf{w}(t), t) \quad (8)$$

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t), M(t), \mathbf{v}(t), t) \quad (9)$$

where the dynamics of the process and the measurements depend not only on the state $\mathbf{x}(t)$ and the process and

measurement noise $\mathbf{w}(t)$ and $\mathbf{v}(t)$ at time t, but also on the model $M(t)$ that governs the system at time t. The probability of each of those models being effective at time t is coded in the discrete probability vector $\mu_{\mathbf{k-1|k-1}}$, as explained in section II.

Figure 1 shows graphically the structure of the IMM estimator. The whole algorithm is detailed in Figure 2. The central part of the algorithm consists of a bank of r filters running in parallel, each one under a different model. An overall estimation for the state can be obtained as a sum of the a posteriori estimation of every filter weighted with the discrete a posteriori model probabilities.

A key aspect of the IMM algorithm is the reinitialisation of the filter before the parallel computation of the filter bank at every step. This mixing of the estimations allows individual poor estimates caused by model mismatch to recombine with estimates from better models, so that the whole filter bank benefits from the better estimates.

IV. INTERACTING MULTIPLE MODEL MONOCULAR SLAM ALGORITHM

Given the tracking-oriented IMM algorithm, some aspects have to be taken into account before applying it to our particular monocular SLAM problem.

1) **Active search ellipses:** In the multiple model tracking literature, little attention is given to the matching (data association) proccess, which is crucial in SLAM algorithms. If matching is mentioned, as in [12], it is said that the most general model, that is, the model with the largest covariance, is used to compute the measurement covariance for gating correspondences — the implication is 'always to expect the worst'.
In monocular SLAM, most of the time this weakest search region is unnecesary large, increasing both the computational cost and the risk of obtaining a false match. A more realistic search region can be defined by the combination of the individual filters weighted by their discrete probabilities. The only assumption that has to be made is that motion changes are smooth, a reasonable assumption when dealing with image sequences. The form of the image search regions is therefore determined by the following equations:

$$\hat{\mathbf{x}}_{k|k-1} = \sum_i \hat{\mathbf{x}}^i_{k|k-1}\mu^i_{k|k-1} \quad (10)$$

$$P_{k|k-1} = \sum_i (P^i_{k|k-1} + (\hat{\mathbf{x}}_{k|k-1} - \hat{\mathbf{x}}^i_{k|k-1}) \quad (11)$$

$$(\hat{\mathbf{x}}_{k|k-1} - \hat{\mathbf{x}}^i_{k|k-1})^\top)\mu^i_{k|k-1} \quad (12)$$

$$\mathbf{h}_{k|k-1} = \mathbf{h}(x_{k|k-1}) \quad (13)$$

$$S_k = H_k P_{k|k-1} H_k^\top + R_k \quad (14)$$

2) **Map management:** As detailed in [6], map management strategies for deleting bad features and adding new ones are convenient in monocular SLAM. We are also using inverse depth to cartesian conversion [3] in order to reduce the computational cost of the algorithm.

V. EXPERIMENTAL RESULTS

A 1374 frame sequence was recorded with a 320×240 wide-angle camera at 30fps. The camera makes a motion consisting of the following sequence of essential movements: stationary → pure rotation → general motion (translation and rotation) → pure rotation → stationary. The sequence has been processed using the dimensionless inverse depth formulation of [2] and two different types of motion modelling. Firstly, IMM EKF formulation with a bank of seven models: stationary camera, rotating camera (three angular acceleration levels with standard deviation 0.1, 0.5 and 1 pixels), and general motion (with 3 acceleration levels for both linear and angular components with standard deviations of 0.1, 0.5 and 1 pixels). Secondly, as a base reference, a single model for general motion with acceleration noise standard deviation of 1 pixel, both angular and linear. Both formulations are fed the same starting image feature detections. On analysing the results the advantages of the IMM over single model monocular SLAM become clear. Results of the comparative experiments can be better observed in the accompanying video (`imm.mpg`).

A. Consistent start up even with rotation

As was said in the introduction, single model EKF SLAM leads to inconsistent mapping if the camera initially undergoes low parallax motion. In the analysed sequence, we have an extreme case of this as the camera is either stationary or rotating for more than 600 frames. Figure 3 compares the estimation results with a single model EKF and our IMM algorithm at step 600, when the camera has performed non-translational motion. Features are plotted as arrows if (as should be the case) no finite depth has been estimated after the no parallax motion. It can be observed that, for the single model case, all features have collapsed to narrow, *false*, depth estimates while in the IMM case all of the features have no depth estimation.

B. Low risk of spurious matches due to small search regions

It can be noticed in Figure 4 that although high process noise models are necessary in order to retain tracking features during high accelerations, these models are scarcely used for any length of time. In hand-held camera sequences, constant velocity motions are much more common than accelerated ones. This is reflected by the model probabilities, as we see that the highest probabilities are given to the lower acceleration noise models on most frames.

When using a single model estimation, we are forced to choose the most general model in order to maintain tracking under high acceleration. As process noise directly influences search region size, we are forced to maintain large search regions, unnecessary most of the time. As a consequence, the risk of obtaining false matches grows. As IMM selects at any time the most probable motion model, preferring simpler models, it adjust the search region to the real motion at any time, resulting in considerably reduced ellipses and lowering the risk of mismatches.

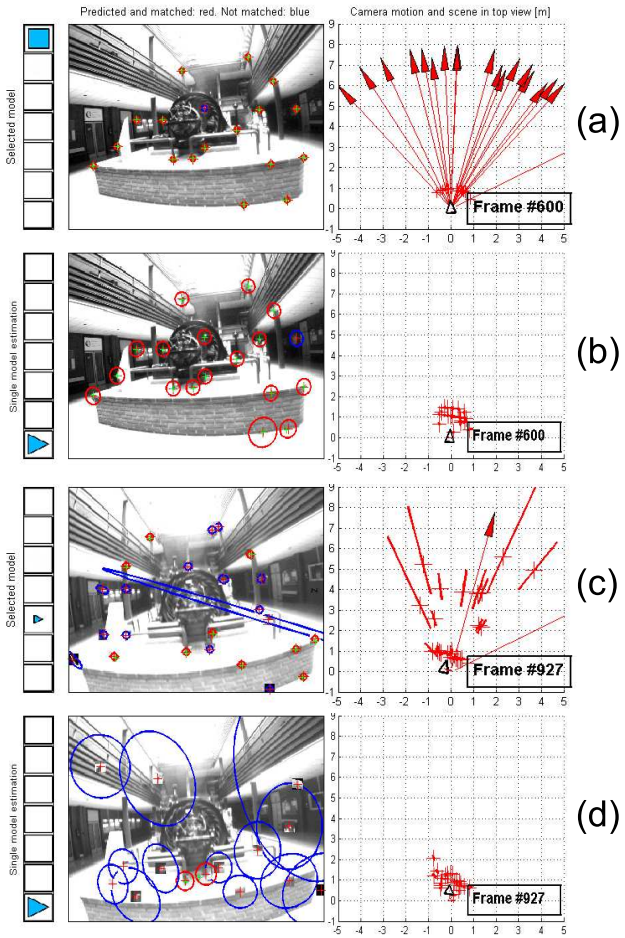

Fig. 3. (a, left) frame 600 and (a, right) 3D top view of the IMM estimation at this frame. The camera has been either stationary or rotating until this frame. It can be seen in Fig. 4 that rotation and still camera models have high probability throughout this early part of the sequence. IMM, correctly, has not estimated any feature depth –features whose depths have not been estimated (their depth uncertainties, stored in inverse depth formulation, encompass infinity) are plotted as arrows–. (b), frame and top-viewed estimation with single-model monocular SLAM. The over-general model has led to narrow, false depth estimates. When the camera translates this inconsistent map leads to false matches that cause the estimation to fail, as seen in (d) at frame 927 of the sequence. On the other hand, (c) shows the correct map estimation performed by the IMM algorithm.

In Figure 5 the large factor of reduction in the size of search ellipses can be observed. Subfigure (a) shows a detail of a feature search region at frame 100, at the top using IMM and at the bottom using a single model. Search regions in subfigure (b) correspond to the same feature at frame 656, when camera starts translating and high acceleration is detected. Notice that the IMM ellipse slightly enlarges in adapting to this motion change, but continues to be smaller than the single-model one. Finally, (c) exhibits the consequences of having unnecessary big search regions: false correspondences happen. Due to mismatches like this one, the estimation in this experiment fails catastrophically.

C. Camera motion model identification

The IMM not only achieves better performance in processing the sequence, but also provides a tool to segment

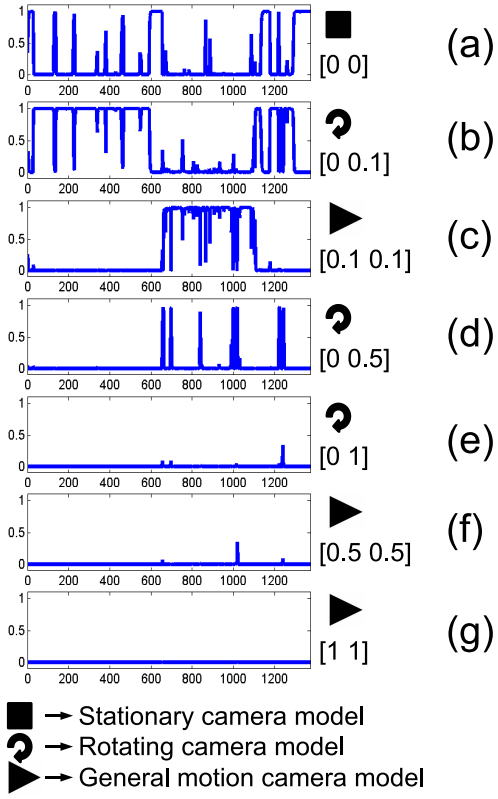

Fig. 4. Posterior model probabilities along the sequence. Each model is represented by its acceleration noise standard deviation$[\sigma_a, \sigma_\alpha]$ expressed in pixels, following the notation in [2]. Notice that the probability for the most general model ($\sigma_a = 1pxl, \sigma_\alpha = 1pxl$) is always under 0.01. The stationary camera model (a) and low acceleration noise models (b) and (c) are assigned the highest probabilities in most of the frames. In spite of being rarely selected, the high acceleration noise models are important to keep the features track at the frames where motion change occurs (small spikes are visible at these points).

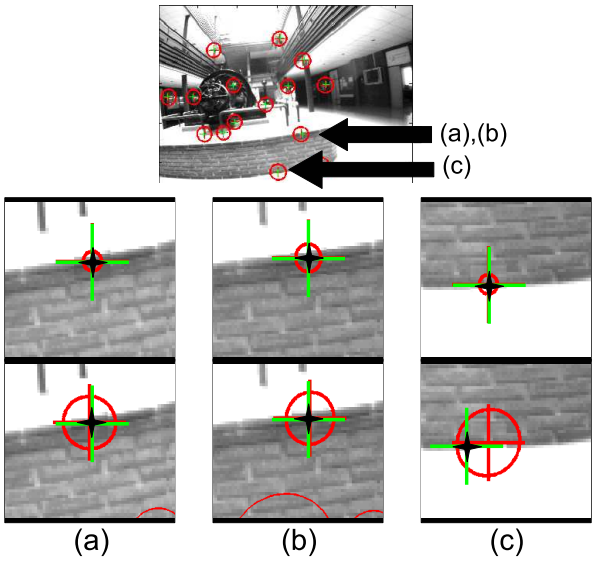

Fig. 5. (a), IMM (top) and single-model (bottom) feature search ellipse when the camera is rotating. (b), the same feature IMM and single-model search regions when the camera begins to translate. (c), mismatch in the single-model case caused by an unnecesary large ellipse that does not occur in the IMM estimation. Several mismatches like this one in the highly repetitive texture of the brick wall eventually lead to full tracking failure.

the sequence according to the dominant motion model. It is worth noting that this segmentation is based on sequence criteria as opposed to a classical pairwise motion model selection in geometrical vision.

In Figure 4 and in video `imm.mpg`, it can be seen that when there is a predominant model (stationary, rotating or general motion), the corresponding probability μ^i reaches a value close to 1, while the other model probabilities goes down close to zero — the IMM acts as a discrete selector here rather than a mixer. Only when there is a change between motion models are there periods with no clear dominant model and this is where the IMM proves its worth.

It has to be noted that models with lower acceleration noise are preferred unless the camera really undergoes a high acceleration motion. In fact the model with the highest acceleration has negligible probability indicating that it is not necessary for processing the current sequence. Although this unused model does require a computational overhead, its presence does not affect the accuracy of the solution nor jeopardize the matching by the size of the search regions for the predicted features — since its weight is always close to zero it is simply weighted out of all the calculations.

D. Computational cost considerations

Although the main advantage of the algorithm is its good tracking performance, clearly outperforming standard single model SLAM on complex sequences, it is also remarkable that the computational cost does not grow excessively. The cost of the IMM algorithm is essentially proportionally to the number of models since all filtering operations must be duplicated for each model. This is offset somewhat, as shown in section V-B, by the fact that the search region ellipses are reduced in size in the IMM formulation and this makes the image processing work of feature matching cheaper.

VI. Conclusions and future work

We have shown experimentally the advantages of the IMM filter when applied to monocular SLAM. We are able to track sequences containing periods with no movement, and pure rotation and general motion at various dynamic levels, the system adapting automatically. In particular, while single model monocular SLAM is weak when bootstrapped with low parallax motions (still or rotating camera), the IMM formulation copes admirably by recognising the motion type.

The IMM formulation requires a computational overhead, but has extra benefits in producing smaller acceptance regions for the predicted measurements, improving outlier rejection, and being able to act as an automatic segmentation and labelling tool by identifying motion boundaries.

This is our first step in the promising direction of Bayesian multiple modelling for monocular SLAM. An immediate next step is to validate empirically the real-time performance of the IMM with an efficient C++ version where performance limits can be more thoroughly tested. Also, and extending the Bayesian approach, the currently fixed transition probabilities between models could be learned and changed dynamically.

VII. Acknowledgments

This work has been partly supported by the MEC DPI2006-13578, MEC PR2007-0427, MEC FIT 360005-2007-9, EPSRC GR/T24685/01, DGA(CONAI+D)-CAI IT12-06 and Royal Society International Joint Project (between University of Oxford, University of Zaragoza and Imperial College) grants. We are grateful for discussions with Simon Julier and Ian Reid.

References

[1] H. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33(8):780–783, August 1988.

[2] J. Civera, A. J. Davison, and J. M. M. Montiel. Dimensionless monocular SLAM. In *Proceedings of the Iberian Conference on Pattern Recognition and Image Analysis*, 2007.

[3] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth to depth conversion for monocular slam. In *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, 2007.

[4] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth parametrization for monocular SLAM. *IEEE Trans. Robotics*, October 2008. Accepted for publication.

[5] L. A. Clemente, A. J. Davison, I. D. Reid, J. Neira, and J. D. Tardos. Mapping large loops with a single hand-held camera. In *Proceedings of Robotics: Science and Systems*, 2007.

[6] A. Davison, I. Reid, N. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.

[7] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the 9th International Conference on Computer Vision, Nice*, 2003.

[8] E. Eade and T. Drummond. Scalable monocular SLAM. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York*, 2006.

[9] A. W. Fitzgibbon and A. Zisserman. Automatic camera recovery for closed or open image sequences. In *Proc. European Conference on Computer Vision*, pages 311–326. Springer-Verlag, June 1998.

[10] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[11] K. Kanatani. Uncertainty modeling and model selection for geometric inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1307–1319, 2004.

[12] T. Kirubarajan, Y. Bar-Shalom, W. Blair, and G. A. Watson. IMMPDAF for radar management and tracking benchmark with ECM. *IEEE Transactions on Aerospace and Electronic Systems*, 34(4):1115–1134, October 1998.

[13] X. R. Li and V. P. Jilkov. Survey of maneuvering target tracking. part v: Multiple-model methods. *IEEE Transactions on Aerospace and Electronic Systems*, 41(4):1255–1320, October 2005.

[14] J. M. M. Montiel, J. Civera, and A. J. Davison. Unified inverse depth parametrization for monocular SLAM. In *Proceedings of Robotics: Science and Systems, Philadelphia*, 2006.

[15] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd. Real-time localization and 3D reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[16] J. Neira and J. D. Tardós. Data association in stochastic mapping using the joint compatibility test. *IEEE Trans. Robotics and Automation*, 17(6):890–897, 2001.

[17] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004.

[18] K. Schindler and D. Suter. Two-view multibody structure-and-motion with outliers through model selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):983–995, 2006.

[19] J. Solà, M. Devy, A. Monin, and T. Lemaire. Undelayed initialization in bearing only SLAM. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems*, 2005.

[20] P. Torr, A. Fitzgibbon, and A. Zisserman. The problem of degeneracy in structure and motion recovery from uncalibrated image sequences. *IJCV*, 32(1):27–45, 1999. Marr Prize Paper ICCV 1999.

[21] P. H. S. Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision*, 50(1):35–61, 2002.