Abstract of the PhD Thesis

Real-Time EKF-Based Structure from Motion

Defended by

Javier CIVERA

on September 9th, 2009

Advisor: Dr. José María MARTÍNEZ MONTIEL

Robotics, Perception and Real Time Group

Aragón Institute of Engineering Research (I3A)

Universidad de Zaragoza, Zaragoza, Spain.

Jury :

Reviewers :	Joan SOLÀ	-	LAAS, Toulouse, France.
	Gabe SIBLEY	-	Oxford University, United Kingdom.
President :	Wolfram BURGARD	-	Universitat Freiburg, Germany.
Examinators :	Juan Domingo TARDÓS	-	Universidad de Zaragoza, Spain.
	Patric JENSFELT	-	KTH, Stockholm, Sweeden.
	Miguel Ángel SALICHS	-	Universidad Carlos III, Madrid, Spain.
	Víctor F. MUÑOZ	-	Universidad de Málaga, Spain.



One of the most brilliant quotes attributed to Albert Einstein says that you do not really understand something unless you can explain it to your grandmother. With that in mind, I could consider myself rather happy, being able to summarize the main aim of my thesis with a simple and understandable sentence like "making a robot see". On the other hand, the lexical simplicity of this objective hides a very complex reality which very often people are tricked into. Even relevant researchers of the field are said to have fallen into the trap: The anecdote that Marvin Minsky, Artificial Intelligence pioneer from MIT, assigned to solve the computer vision problem as a summer project to a degree student back in the sixties is an illustrative and well-known example [26].

The truth behind this apparent simplicity is that, although we all have a clear experience about what "to see" implies, the biological mechanisms of visual processing are still not fully understood. And even if we knew it, we could also wonder if a machine needs –or will be able to run– a visual sensing similar to ours. This lack of a precise definition about what "to see" really means and needs in an algorithmic sense have made of Computer Vision a diverse and fragmented discipline.

In spite of this, Computer Vision has experienced great advances since its appearance. Computers still cannot see, but most of them nowadays use visual information in one or another sense. And mainly because of the richness of the visual information, cameras are nowadays the dominant sensor in the Robotics research. One of the uses of visual information is the general frame of this thesis, specifically how visual information can be processed to extract a tridimensional estimation of the imaged scenario and the motion of the camera into it.

The algorithms described in this thesis provide a theoretical framework to perform 3D estimation of a camera motion and 3D scene from the only input of an image sequence and in real-time up to 30 frames per second. Compared with the state-of-the-art on the topic, the algorithms described allow for the first time to perform 3D estimation *out-of-the-box*; that is, for any sequence and any camera motion, assuming no knowledge over the scene nor the internal camera calibration parameters.

This comes from the application of solid theoretical concepts, deeply rooted in Projective Geometry and Probability. As another output of the application of a well-founded theory, the length of the experimental results greatly increases: state-of-the-art experiments are limited to waggling camera motion in indoors scenarios and sequences of around a minute. Here, sequences of tens of thousands frames taken by a robot covering trajectories of hundreds of metres in about half an hour serve as input for highly accurate camera motion estimation.

I. State-of-the-Art

A. Pairwise Structure from Motion

Inside the Computer Vision field, Structure from Motion (SfM) is the line of research that takes as input the 2D motion from images and seeks to infer in a totally automated manner the 3D structure of the scene viewed and the camera locations where the images where captured. SfM has been a very active area of research for the latest three decades, reaching such a state of maturity that some of its algorithms have already climbed to the commercial application level [1], [2], [33].

SfM origins can be traced back to the so-called Photogrammetry, that since the second half of $19th$ century aimed to extract geometric information from images. Starting with a set of features manually identified by the user, Photogrammetry make use of non-linear optimization techniques known as Bundle Adjustment (BA) [30]. Computer Vision research has been oriented to achieve the complete automation of the problem, producing remarkable progress in two aspects: first, the constraints imposed on the motion of the features in two images under the assumption of the rigidity of the scene have been formalized [26]. And second, intense research on salient feature –points or lines– detection and description [7], [25], [27] and spurious rejection [23] has provided with an automated way of robustly extracting and matching those salient features along images. Imposing the algebraic constraints over the matched salient features provides the equations to estimate the camera motion and 3D feature location.

Based on these two achievements, several methods have been proposed that, from the only input of a set of images from a scene can estimate the tridimensional structure and pairwise relative camera motion up to a projective transformation in the most general case of uncalibrated cameras. With some extra knowledge about camera calibration, a solution up to scale can be obtained. In most cases, the SfM pairwise initial estimation is

used as a the initial seed in a global or local Bundle Adjustment [40] –as Photogrammetry does– that refines the initial estimation into a more accurate and globally consistent one.

B. Filtering-Based Structure from Motion

The key difference of filtering methods from the pairwise ones describe above is that they do not operate in a relative manner –estimating motion from one image with respect to another one– nor do they pile up correspondences waiting for a local BA optimization. Instead of that, the overall state of the system is summarized into a multidimensional probability distribution, and measurements are processed and their information integrated in this probability distribution sequentially as they are gathered. Therefore, its computational complexity scales with the size of the state and not with the number of frames, being naturally suited for the processing of video sequences.

Filtering methods for visual 3D estimation has been a classical line of research within the Computer Vision community. Early work began to appear in the $90s$ [28], [5] using an Extended Kalman Filter to estimate 3D structure and camera motion from the correspondences in a monocular sequence. Results from this early research were far from satisfactory due to a deficient modeling of the problem, that has been refined up to the present: In [3], [35] partial self-calibration was included and in [10] observability was analyzed and occlusions were better modeled.

Filtering methods for 3D estimation from sensor information have been extensively used and developed in the robotic SLAM (Simultaneous Localization and Mapping). The main reason for that is the need for sequential algorithms in robotic applications. The main objective of Computer Vision is to produce the best possible results from a set of images without time constraints. The control loop in robotic systems need sequential and efficient online estimation that provides the estimation results up to the present step with minimum delay.

In general terms, SLAM seeks to estimate a map of the environment and the location of the robot inside of it from the information gathered by sensors attached to the robot. In the most typical SLAM problem, sensory information comes from proprioceptive sensors –odometry or inertial measurement units– and exteroceptive sensors, that measure entities external to the robot. Traditionally, laser has been the predominant exteroceptive sensor used in SLAM, although other sensors, like sonars, have also been used. The Extended Kalman Filter may be the most popular algorithm and the first one to provide a sequential solution to the problem [37], [21], [8], [39].

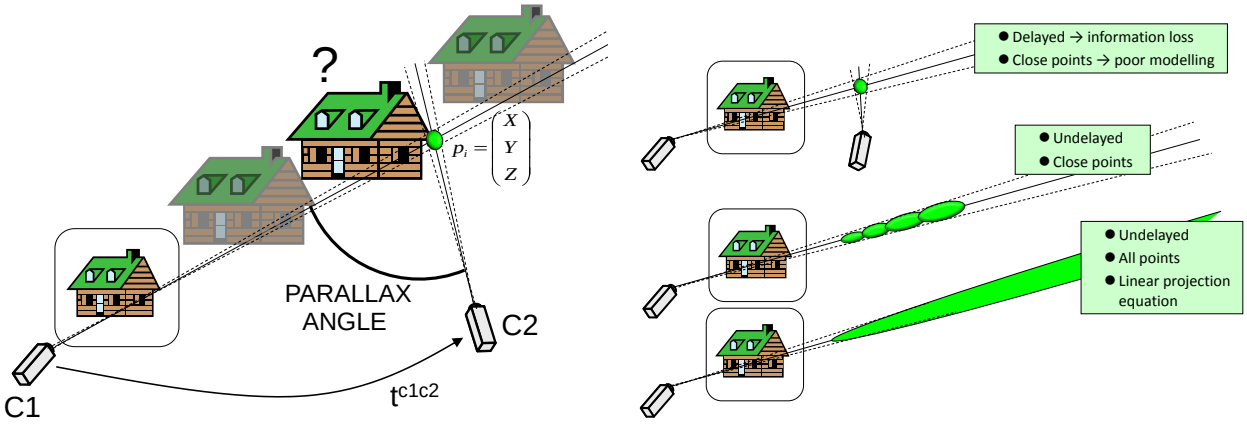
Andrew Davison's seminal work [18], [19] represents the state of the art at the beginning of this thesis and the meeting point where Computer Vision and Robotics research converge being cameras adopted as the main sensor in SLAM. Its main contributions are a better model for an image sequence and an efficient correspondence search algorithm. As a result, in this work real-time visual estimation at 30 frames per second is achieved, but still with severe constraints: it is required an indoors environment and initial camera translation, small sequences of around a minute long can only be processed as a result of not considering spurious measurements and camera has to be precalibrated. Taking Davison research as starting point, this thesis achieves a total relaxation of the constraints over filter-based visual estimation and a great improvement over previous results. The specific contributions of the thesis are briefly detailed in next section.

II. Contributions of this thesis

The main aim of this thesis can be summarized as "to develop models and methods fitting the projective nature of the camera and a Bayesian filtering framework". Specifically, the contributions in this thesis are:

A. Inverse Depth Parameterization for Point Features [31], [14], [15].

This point model definitely closes a large stream of research going from early research within the computer vision community [3] on how to represent point features in filtering-based visual estimation; up to the very recent efforts, mainly from the robotics community, to solve the initialization problem [4], [38], [22]. The two key contributions in this parametrization are: First, and differently from Euclidean, it is a projective model able to deal with distant –even infinite– and close points in a unified manner. And second, differently from Homogeneous, it is linear enough to hold the tight linearity constraints of EKF filtering.



(a) Point Initialization from a Monocular Camera. From a single image (C1), only the ray where the point feature lives can be recovered, being the depth unknown. Only when camera translates up to C2 and enough parallax is gathered, the point depth can be estimated by triangulation.

(b) Point initialization algorithms. In the first row, point initialization is delayed until enough translation is performed, which causes information loss and inability to model distant point features [4]. In the second row, several hypothesis are deployed along the ray to achieve undelayed initialization [38]. Finally, our proposed inverse depth [15] is able to code in a single hypothesis the uncertainty along the ray up to infinity in a unified and undelayed manner.

Fig. 1. Point Initialization and Inverse Depth Coding. (a) illustrates the need of at least two images to estimate the depth of a point from a monocular camera. (b) shows the inverse depth coding proposed in this thesis.

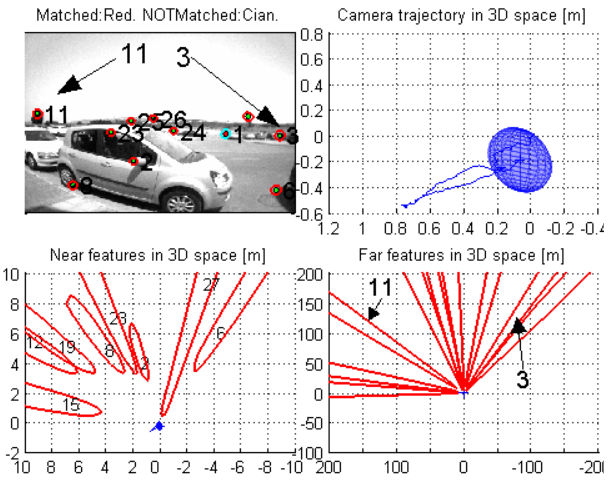
The proposed parametrization can be better understood looking at figure 1. Figure 1(a) illustrates the initialization problem for bearing-only sensors like a monocular camera: If a point feature has to be estimated from a single image C1, the image point can be back projected and the ray where the point lives can be accurately determined. The depth of the point along the ray remains unknown until the camera translates enough to triangulate.

Figure 1(b) illustrates the previous work on point feature initialization within EKF monocular SLAM. The first method used [4] consisted on delay the initialization until an accurate depth could be estimated, resulting in a loss of information and distant –low parallax– features that could never be estimated. An improvement over that was to consider several hypothesis –green ellipses in figure 1(b)– along the ray [38]; allowing undelayed initialization but still not being able to adequately represent distant features. Finally, the explicit estimation of the inverse of the depth along the ray proposed in this thesis allowed to code in a single hypothesis the unknown depth along the ray.

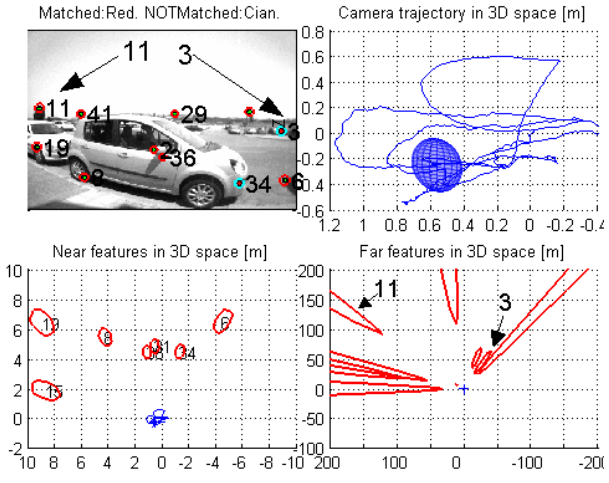
Figure 2 details a real image experiment illustrating the benefit in our proposal. In subfigure 2(a) it is seen an image of the processed sequence and 3D estimation plots at three different scales. Notice the two selected features named 3 and 11: they have recently been initialized and, as a result, do not have an accurate depth estimation: the uncertainty regions, plotted as red lines, expands along the ray. Look now at subfigure 2(b): the camera has translated some metres and now an accurate depth has been estimated for feature 3 –red ellipse represents the estimation uncertainty–, located in a close car. For feature 11, located in a distant tree, the location of the ray is accurately estimated, being a rich source for estimating camera orientation, but depth still cannot be estimated.

B. Efficient 1-Point RANSAC EKF for visual estimation [16], [17].

In any visual algorithm, spurious correspondences are likely to appear and only one of them can spoil the estimation. The proposed 1-point RANSAC EKF is an novel filtering scheme able to detect those spurious matches with a slight cost overhead always lower than 10%. This algorithm has been proved to overcome the Joint Compatibility Branch and Bound [32], gold-standard technique for spuriuos rejection in filtering, both in performance and cost.



(a) Early frame in the sequence: Camera have not translated enough; features 3 and 11 have been initialized without delay and then offer valuable information –mainly about camera rotation– to the estimation. As they have been recently initialized, depth estimation is still rather inaccurate.



(b) Late frame in the sequence: Camera have translated some metres; feature 3 has an accurate estimated depth, feature 11 keeps a consistent large uncertainty in its depth as camera still has not translated enough –but its direction is accurately estimated.

Fig. 2. Inverse Depth EKF Estimation from a Monocular Sequence.

The combination of this spurious rejection, the inverse depth described before and robocentric filtering [9] –used here for the first time in visual estimation– rises two orders of magnitude the allowable camera motion using plain visual filtering: from the waggling camera motion in the seminal paper [19] and predecessors to hundreds of metres trajectories.

Figure 3 shows the achievable accuracy using the proposed algorithm. In this experiment, a mobile robot equipped with a monocular camera covered an outdoors trajectory of around 650 metres, while recording an image sequence of around $24,000$ frames. This image sequence served as only input to the proposed EKF filter. The estimation accuracy can be visually assessed from the figure. The estimation error remains at 1% of the trajectory.

C. Visual Filtering from Uncalibrated Sequences [11].

Camera self-calibration is the process of estimating the internal parameters of a camera from a set of arbitrary uncalibrated images of a general scene. Self-calibration is always preferable and sometimes essential when visual estimation faces real world applications. First, it avoids the onerous task of taking pictures of the calibration object and using a calibration software. Such a task may be difficult for an unexperienced end-user and may be impossible in certain robotic applications; for example if a camera on a robotic arm is not accessible. Second, internal parameters of a camera may change either unintentionally (e.g. due to vibrations, thermical or mechanical shocks) or even intentionally in the case of a zooming camera. 3D estimation in this latter case could only be performed via self-calibration. Finally, inaccurate calibration (coming either from a poor calibration process or from changed calibration parameters) produces the undesirable effect of introducing bias in the estimation.

SfM methods in the decade of the 90's were developed to estimate 3D structure from uncalibrated images, that is, images taken from cameras with unknown and possibly varying calibration parameters. It was also theoretically formalized under what conditions uncalibrated parameters could be estimated, stating the basis for camera self-calibration. Literature over self-calibration using filtering is rather long (e.g. [3], [29], [34]), but any of the presented approaches achieves a complete calibration without prior knowledge. A filtering algorithm is proposed in this thesis that, for the first time, achieves a complete camera calibration from scratch.

Figure 4 details the results offered by the proposed algorithm applied over an uncalibrated sequence. Figure

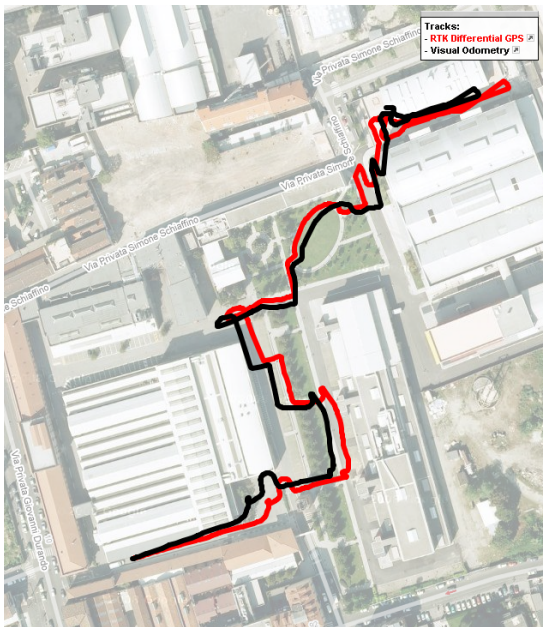


Fig. 3. Estimation results for a robot trajectory of 650 metres from the only input of a monocular sequence taken by a camera mounted in the robot. Red line stands for GPS, the black line is the EKF estimation result. The estimation error is less than 1% of the trajectory.

4(a) details the camera motion and 3D scene estimation results for an early and a late frame of the sequence. Figure 4(b) shows the estimated calibration parameters compared with the ground truth values.

D. Visual Filtering from Degenerate Camera Motion [13].

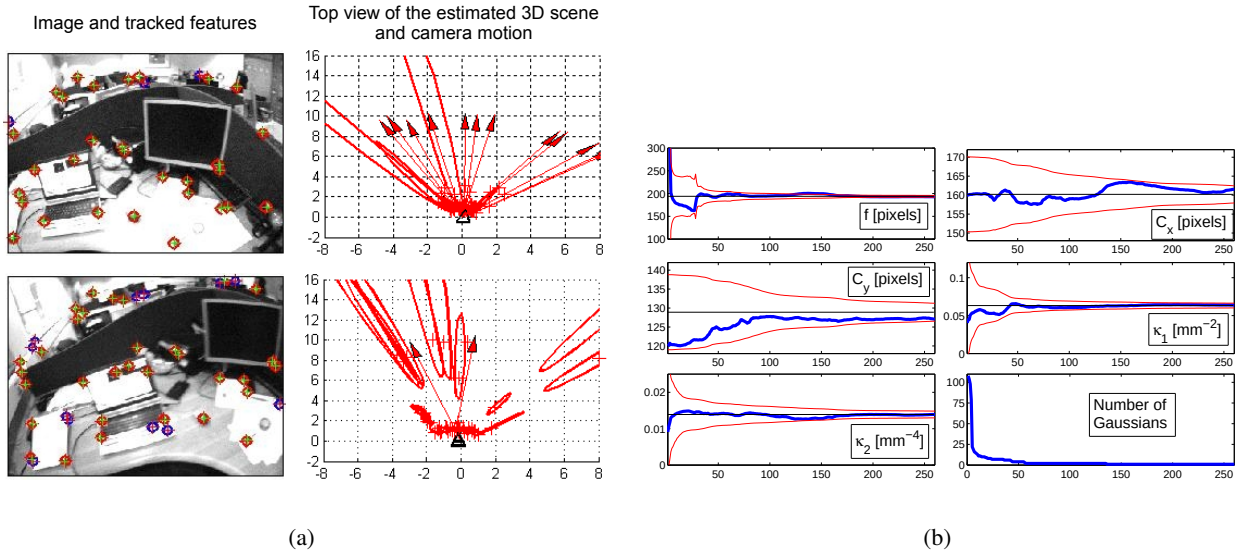
Degenerate camera motions have a capital importance from a practical point of view of implementing real systems. If the camera is attached to a mobile robot, there will occur frequently that the robot is going to be stopped. Pure rotation motion is also very frequent in industrial robotic arms. Any estimation algorithm modeling a general camera motion in any of these situations will fail. In this thesis, it is proposed a model selection scheme well-suited to Bayesian filtering able to produce an accurate and consistent estimation under degenerate motion and seamlessly cope with motion transitions.

The proposed algorithm was tested using real imagery. Figure 5 shows three frames of a sequence performing stationary-rotating-general motion, respectively. At each step, the real motion model was correctly selected: the stationary camera model, represented by a blue square, in (a); the rotating model, represented by a circular arrow in (b); and the general motion, represented by a blue triangle, in (c). The appropriate camera motion model was then applied at each step of the sequence, keeping a consistent estimation of the 3D motion and point features.

E. Drift-Free Real-Time Mosaicing [12].

Mosaicing is a usual application in Computer Vision consisting of stitching together data from a number of images, from a rotating camera or from a translating camera observing a plane, in order to create a composite image which covers a larger field of view than the individual views. Traditional mosaicing algorithms, like [6], based on the application of pairwise SfM algorithms described in section I-A are usually limited in computational time or accuracy. If real-time performance is pursued, then a global optimization cannot be applied and drift accumulates. When drift is to be minimized, an expensive Bundle Adjustment is needed and real-time is lost.

An EKF-based mosaicing algorithm has been developed in this thesis that stitches image textures over the map of feature directions. This mosaicing algorithm directly inherits the EKF advantages of summing up the measurements up to the current step in a state vector, being its cost independent of the number of images.The presented algorithm is the first mosaicing technique that presents real-time drift-free spherical mosaics of 360°. A spherical mosaic constructed in real-time from a sequence gathered by a rotating camera is shown in figure 6.



(a) (b)

Fig. 4. Self-calibration, camera trajectory and 3D scene estimation from an uncalibrated image sequence. (a) shows the amera trajectory and 3D point estimation. (b) shows the estimation results for the calibration parameters. Black lines are the known ground truth values, blue thick lines are the estimated self-calibration parameters and red thin lines the estimated uncertainty. Notice that at the beginning of the sequence uncertainty is large, as we do not have prior knowledge about the parameters. As the estimation evolves, the estimated values (in thick blue) become close to the ground truth values (in black) and uncertainty (thin red) sharply decrease.

III. DISCUSSION

In their short life, the algorithms proposed in this thesis have already reached the status of "standard" and have produced a qualitative jump in visual filtering: Using the state-of-the-art before this research, it could only be estimate waggling camera motion within strongly constrained scenes –containing only close points or needing known objects to initialize–, non-degenerate camera motions and pre-calibrated cameras. In this thesis, visual estimation is performed out-of-the-box, for *any scene* containing close and distant points and for *any sequence*, that means any camera motion and unknown camera calibration.

The algorithms proposed in this thesis are widely used throughout the world. The publications of this thesis have already received more than 250 references from academic documents, according to GoogleScholar. Some of those publications are among the most referenced papers in their respective years, not only in Robotics but in the whole Computer Science area according to Citeseer. Demonstrative code released for testing purposes, either [36] or the most recent one [20] under GNU GPL License, are accessed and downloaded several times per day.

The degree of maturity of this research –and geometric vision in general– and the development of demonstrative real-time software offers, apart from interesting research lines, exciting opportunities for technological transfer. Among the research lines opened by the presented research, two of them stand out as greatly beneficial:

- **Surgeon assistance via augmented reality in endoscopic operations.** Being able to construct a 3D model of an imaged scene implies being able to make measurements of the scene and make virtual insertions. Both capabilities are seen as extremely useful by endoscopic surgeons. First, making measurements allow to know the size of certain body structures, detecting anomalies and re-planning the surgery. Currently, endoscopic measurements are performed by introducing a metric tape inside the body. And second, inserting virtual objects allow to easily limit safe operation areas. Figure 7 shows some preliminary results of this research.
- **Visual sensing for service robots.** My research group is currently involved in the RoboEarth project [41], funded by the EU. The aim of the project is to construct a web-like database for robots to share knowledge: a robot that successfully performs a task can upload a "recipe" to the database that another robot can download and use when required to perform the same action. The group is required to construct the shareable visual sensing capabilities of the robots accessing the database.

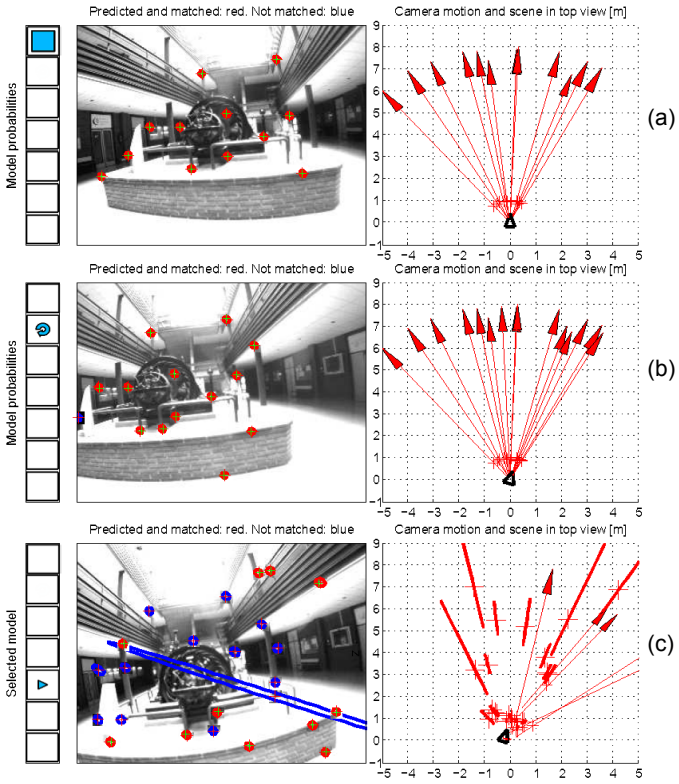


Fig. 5. Model Selection Results: When a degenerate camera motion –as stationary camera in (a) or rotating camera in (b)–, the stationary model and rotating model are correctly detected and estimation evolves assuming degenerate models. When the camera finally translates

Algorithms developed in this thesis are essential in the above described lines: the highest degree of robustness and real-time are critic for endoscopic surgery. Self-calibration is also needed to alleviate the time needed to prepare the system. Real-time and robust operation are also a must for real robots. Both lines involve challenging research: highly dynamic and non-rigid environments in the endoscopic case and recognition capabilities for service robots.

Regarding technological transfer, the presented research is currently being implemented for commercial use and protected via patents in three different applications with three different companies: 1) a robust indoor localization software; 2) a grasping system for a robotic arm; and 3) real-time virtual insertions for broadcasting of sports events.

REFERENCES

[1] 2d3. URL http://www.2d3.com/, May 2009.

[2] Autosticht. URL http://www.autostitch.net/, May 2009.

[3] A. Azarbayejani and A. P. Pentland. Recursive estimation of motion, structure, and focal length. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6):562–575, June 1995.

[4] T. Bailey. Constrained initialisation for bearing-only SLAM. In *Proc. IEEE Int. Conf. Robotics and Automation*, Taiwan, 2003.

[5] T. Broida, S. Chandrashekhar, and R. Chellappa. Recursive 3-d motion estimation from a monocular image sequence. *IEEE Transactions on Aerospace and Electronic Systems*, 26(4):639–656, 1990.

[6] D. Brown, M.and Lowe. Recognising panoramas. In *International Conference on Computer Vision*, pages 1218–1225, Nice, 2003.

[7] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[8] J. Castellanos, J. Montiel, J. Neira, and J. Tardos. The SPmap: a probabilistic framework for simultaneous localizationand map building. *IEEE Transactions on Robotics and Automation*, 15(5):948–952, 1999.

[9] J. Castellanos, J. Neira, and J. Tardos. Limits to the consistency of EKF-based SLAM. In *5th IFAC Symposium on Intelligent Autonomous Vehicles*, July 2004.

[10] A. Chiuso, P. Favaro, H. Jin, and S. Soatto. "MFm": 3-D motion from 2-D motion causally integrated over time. In *European Conference on Computer Vision*, pages 735–750, 2000.

[11] J. Civera, D. R. Bueno, A. Davison, and J. Montiel. Camera Self-Calibration for Sequential Bayesian Structure From Motion. In *IEEE International Conference on Robotics and Automation, 2009. ICRA 2009*, 2009.

[12] J. Civera, A. Davison, J. Magallon, and J. Montiel. Drift-Free Real-Time Sequential Mosaicing. *International Journal of Computer Vision*, 81(2):128–137, February 2009.
[13] J. Civera, A. Davison, and J. Montiel. Interacting multiple model monocular SLAM. In *IEEE International Conference on Robotics and Automation, 2008*, pages 3704–3709, 2008.
[14] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth to depth conversion for monocular SLAM. In *IEEE International Conference on Robotics and Automation, 2007*, pages 2778–2783, April 2007.
[15] J. Civera, A. J. Davison, and J. M. M. Montiel. Inverse depth parametrization for monocular SLAM. *IEEE Transactions on Robotics*, 24(5):932–945, October 2008.
[16] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel. 1-point RANSAC for EKF-based structure from motion. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3498–3504, October 2009.
[17] J. Civera, O. G. Grasa, A. J. Davison, and J. M. M. Montiel. 1-point ransac for ekf filtering: Application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, October 2010. to appear.
[18] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Ninth IEEE International Conference on Computer Vision, 2003. Proceedings*, pages 1403–1410, 2003.
[19] A. J. Davison, N. D. Molton, I. D. Reid, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, June:1052–1067, 2007.
[20] .-P. R. E. M. demonstrative code. http://webdiis.unizar.es/ jcivera/code/1p-ransac-ekf-monoslam.html, May 2010.
[21] M. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba. A solution to the simultaneous localization and map building (SLAM) problem. *IEEE Transactions on Robotics and Automation*, 17(3):229–241, 2001.
[22] E. Eade and T. Drummond. Scalable monocular SLAM. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York*, pages 469–476, 2006.
[23] M. A. Fischler and R. C. Bolles. Random sample consensus, a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381 – 395, 1981.
[24] O. G. Grasa, J. Civera, A. Guemes, V. Muoz, and J. M. M. Montiel. Ekf monocular slam 3d modeling, measuring and augmented reality from endoscope image sequences. In *5th Workshop on Augmented Environments for Medical Imaging including Augmented Reality in Computer-Aided Surgery, held in conjunction with MICCAI2009*, 2009.
[25] C. G. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
[26] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, 2004.
[27] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
[28] L. Matthies, T. Kanade, and R. Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3):209–238, 1989.
[29] P. McLauchlan and D. Murray. Active camera calibration for a head-eye platform using a variable state-dimension filter. Accepted for PAMI, 1994.
[30] E. Mikhail, J. Bethel, and M. J.C. *Introduction to Modern Photogrammetry*. John Wiley & Sons, 2001.
[31] 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, USA, August 2006.
[32] J. Neira and J. D. Tardós. Data association in stochastic mapping using the joint compatibility test. *IEEE Transactions on Robotics and Automation*, 17(6):890–897, 2001.
[33] PhotoTourism. URL http://phototour.cs.washington.edu/, May 2009.
[34] G. Qian and R. Chellappa. Bayesian self-calibration of a moving camera. *Computer Vision and Image Understanding*, 95(3):287–316, 2004.
[35] G. Qian and R. Chellappa. Structure from motion using sequential monte carlo methods. *International Journal of Computer Vision*, 59(1):5–31, 2004.
[36] . S. S. School. URL http://www.robots.ox.ac.uk/ SSS06/Website/Practicals.htm, May 2010.
[37] R. Smith, M. Self, and P. Cheeseman. A stochastic map for uncertain spatial relationships. In *4th International Symposium on Robotics Research*, 1987.
[38] J. Sol, M. Devy, A. Monin, and T. Lemaire. Undelayed initialization in bearing only SLAM. In *IROS*, 2005.
[39] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. Cambridge: MIT Press, 2005.
[40] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment – A modern synthesis. In *Vision Algorithms: Theory and Practice*, LNCS, pages 298–375. Springer Verlag, 2000.
[41] R. Web. http://www.roboearth.org/, May 2010.

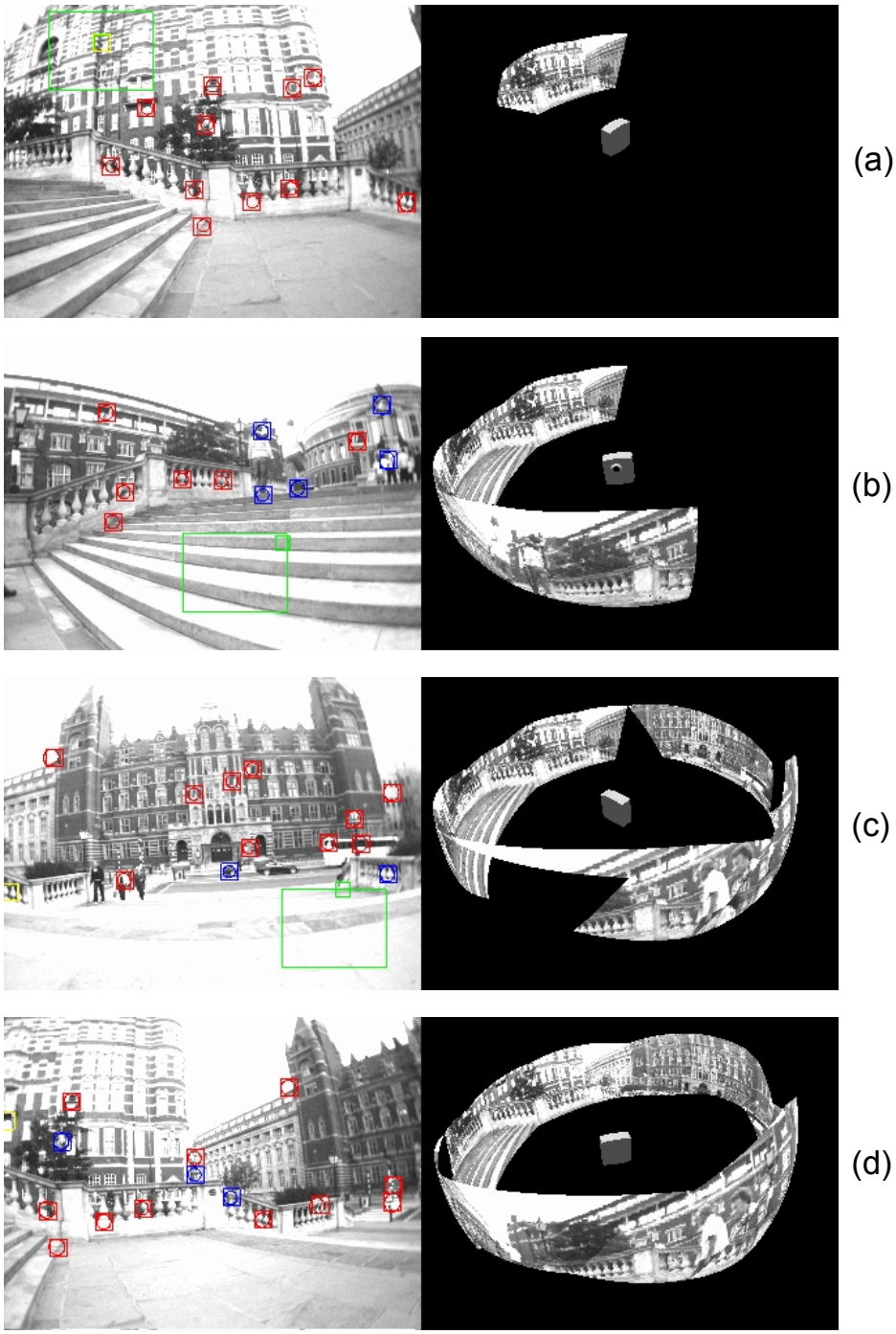


Fig. 6. Drift-Free Real-Time Online Mosaicing Results. (a) shows an early frame of the sequence and the initial mosaic. (b) shows a 180° mosaic in a more advanced frame of the sequence. (c) shows the recognition of the starting point and 360° loop closing. (d) shows the final mosaic. All the processing was done in real-time at 30 frames per second.

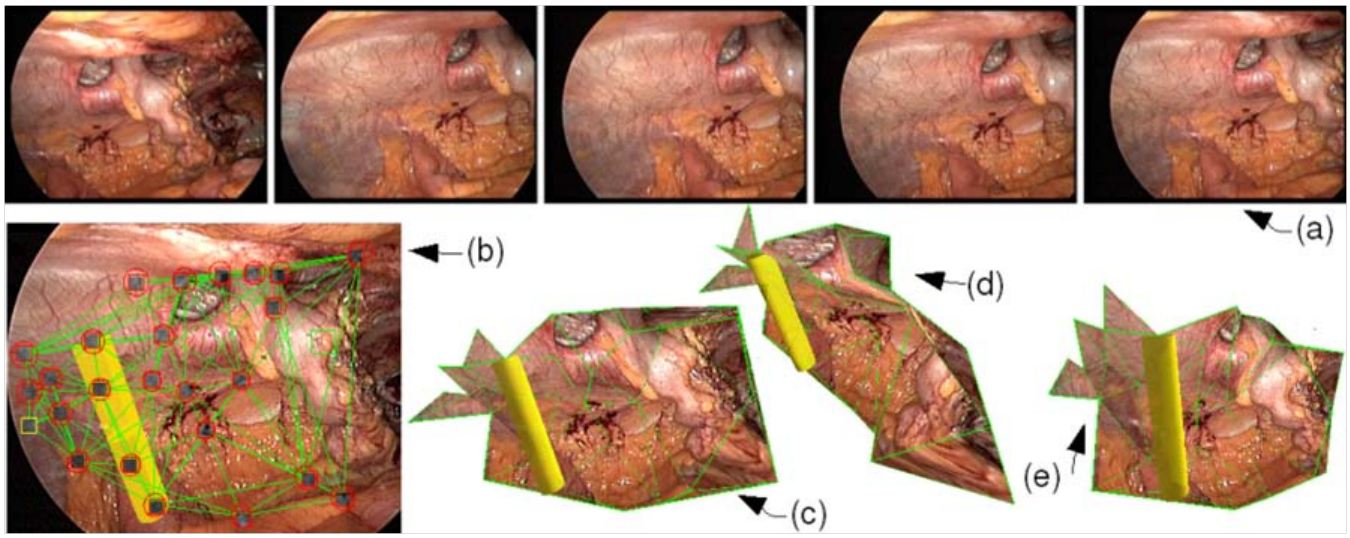


Fig. 7. New research in visual 3D estimation applied to endoscopic surgery. This early work has been published in [24]. (a) shows several frames taken by an endoscopic camera. The real-time demonstrative software developed in this thesis is applied over the sequence (b), producing the 3D models shown in (c), (d) and (e). Such 3D models are used to help the surgeons with measurement capabilities and augmented reality insertions –like the yellow cylinder in the figures– to delimit safe areas.