Hybrid metric-topological mapping for large scale monocular SLAM

Eduardo Fernández-Moral, Vicente Arévalo and Javier González-Jiménez

MAPIR Group. Universidad de Málaga. E.T.S. Ingeniería de Informática-Telecomunicación. Campus de Teatinos, 29071 Málaga, Spain
{eduardofernandez,varevalo,javiergonzalez}@uma.es
http://mapir.isa.uma.es/

Abstract. Simultaneous Localization and Mapping (SLAM) is a central problem for autonomous mobile robotics. Monocular SLAM is one of the ways to tackle the problem, where the only input information are the images from a moving camera. Current approaches for this problem have achieved a good balance between accuracy and density of the map, however, they are not suited for large scale. In this paper, we present a dynamic mapping strategy where the metric map is divided into regions with highly connected observations, resulting in a topological structure which permits the efficient augmentation and optimization of the map. For that, a graph representation where the nodes represent keyframes, and their connections are a measure of their overlapping, is continuously rearranged. The experiments show that this hybrid metric-topological approach outperforms the efficiency and scalability of previous approaches.

Keywords: Monocular SLAM, metric-topological mapping, map partitioning

1 Introduction

Monocular SLAM is an appealing way of solving the localization and mapping problem in mobile robotics because cameras are inexpensive, compact, easy to calibrate and consume low power. During the last years monocular SLAM has advanced notably with the use of parallel processing and efficient algorithms for data association and map optimization. It has made possible that current state-of-the-art approaches can operate accurately in some large scale scenarios, facilitating its application in a wide range of areas such as augmented reality, scene reconstruction and, particularly, mobile robotics.

The increasingly larger maps that are now possible with monocular SLAM are fundamental to cope with a wider range of real autonomous robotics applications. Such ability to operate in large scale brings the need of appropriate strategies for managing the map. Applying abstraction (as humans do) is an effective way of dealing with the huge amount of detail present in large metric maps. The result of such abstraction process is the so-called metric-topological map, consisting of

E. Fernandez-Moral, V. Arevalo, J. Gonzalez-Jimenez, "Hybrid Metric-Topological Mapping for Large Scale Monocular SLAM", Informatics in Control, Automation and Robotics, Lecture Notes in Electrical Engineering, INSTICC, vol. 325, pp. 217-232, 2015.

a two-layer representation, one containing pure geometrical information and a second one containing higher level symbolic information [25].

The benefit of a metric-topological arrangement is twofold: on the one hand, it offers a natural integration with symbolic planning that permits a robot to reason about the world and to execute high level tasks [10]. On the other hand, the efficiency and scalability of the SLAM process itself are improved by limiting the scope of localization and mapping to the region of the environment where the robot is operating. Also, loop closure and relocalisation can be more efficiently solved using topological information [20, 1, 9].

In this work, we present an online submapping technique which creates a topological representation of the world from the metric map being built by a monocular SLAM technique[1]. The key idea of our proposal is to cluster in the same submap those keyframes with higher observation overlap. This presents some important advantages over other approaches (as it will be explained latter on). The generated map consists of a topological structure composed of nodes representing local metric maps and arcs representing relative geometric transformations among the so-called submaps. In this paper, we will focus on the benefits of such a hybrid map for improving the efficiency and scalability of conventional (metric) monocular SLAM, concretely PTAM [12].

Next, we discuss some relevant related work and explain in detail the advantages of our approach. We then describe our partitioning procedure and show how it is combined with the SLAM process (PTAM). The experiments and their results are presented next, and finally, we expose the conclusions of our work.

2 Related Works

2.1 Construction of the Metric Map

Many solutions have been presented to build metric maps with monocular SLAM since Davison [5] presented the first real-time solution for the problem in 2003. Two main strategies have been applied since then: Bayesian filtering (following the work of Davison) and Bundle Adjustment (BA) on keyframes, as introduced in [12]. The latter represents the base for the current state of the art since it allows handling denser maps and generally offers a better ratio accuracy/cost [24].

BA, traditionally used as an offline method for Structure from Motion (SfM), is now widely used in visual SLAM thanks to the introduction of parallel processing and efficient algorithms which exploit the sparse structure of the problem. Its application to visual SLAM was inspired by real time visual odometry and tracking [18], where the most recent camera poses where optimized to achieve accurate localization. In such line, PTAM selects keyframes and applies BA in a fixed size window, around the last keyframe incorporated, to obtain good metric

[1] A preliminary version of this paper was presented in the "10th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Reykjavík (Iceland), 2013" [8].

maps and accurate localization. Then, once the local optimization is performed, a low priority global BA is run to improve the map consistency. This approach is extended in [11] by combining it with relative bundle adjustment -RBA- [22], allowing fixed-time, consistent exploration. An improvement of the latter to exploit the problem' sparse structure was recently presented by [4].

The work of [23] is also related to RBA, they propose a double window optimization: a first window as in PTAM and a second one including the periphery of the first to improve consistency by optimizing a pose-graph. Despite the impressive results obtained, such unique map solution has intrinsic limitations for managing maps of real large environments. To avoid such a limitation, we propose a topological arrangement in local metric maps.

2.2 Dividing the Map

Map division has been addressed in a number of works. Some relevant examples are: the Atlas framework [15], where a new local map is started whenever localization performs poorly in the current local map, or the hierarchical SLAM presented in [7], where sensed features are integrated into the current local map until a given number of them is reached. However, none of these provides a mathematically grounded solution based on the particular perception of the scene.

In [6], the map is divided in nodes where the landmarks are represented in a local coordinate frame and, these landmarks are updated using an information filter. This method uses the common features between adjacent nodes to calculate their relative pose. A different approach called Tectonic-SAM [17] uses a "divide and conquer" approach with locally optimized submaps in a Smoothing and Mapping framework (SAM). This approach is improved in [16] to build a hierarchy of multiple-level submaps using nested dissection.

Other works employ "graph cut" to divide the map according to a measurable property of the map observations. On that mathematical sound basis, [26] addresses the problem of automatic construction of a hierarchical map from images; [2] generates metric-topological maps using a range scanner, and generalizes the approach for other sensors; and [19] splits the map within a Bayesian monocular SLAM framework to reduce the problem complexity.

Our method, which also relies on graph cut, differs from the above works in the way the graph is constructed, which is specifically tailored for BA-based monocular SLAM. Our approach resembles also the stereo-SLAM framework of [14] who divide the map keyframes into groups (called segments) according to their geodesic distances in the graph. On the contrary, our map partitioning is independent of the keyframe positions, and is only based on observations acquired from the scene. Concretely, the map is split where there are less shared observations, minimizing the loss of information and therefore, enforcing the coherency and consistency of the submaps.

3 Map Partitioning

Splitting a map into locally metric consistent and globally coherent regions provides some relevant advantages for SLAM. Next, we explain the benefits of such map structure (subsection 3.1), and describe our proposal to obtain this metric-topological arrangement of the map (subsection 3.2).

3.1 SLAM Improvements through Hybrid Mapping

The advantages of applying a coherent map partition in monocular SLAM are diverse: a) all the metric data in each submap can be referred to a local coordinate system, what reduces error accumulation and numerical instability; b) localization can be achieved more efficiently since only those map points in the nearer regions are reprojected to estimate the camera position; c) this map structure permits to approximate the global BA by the individual optimization of the different submaps, thus reducing the computational cost of the optimization process. This last advantage is of special relevance due to the demanding nature of BA, whose complexity ranges from linear to cubic in the number of keyframes depending on the particular point-keyframe structure [13]. Next, we explain the details of this approximation for the global optimization.

Having a map of n landmarks obtained from observations at m keyframes, bundle adjustment can be expressed as

$$\min_{\mathbf{a}_j,\,\mathbf{b}_i} \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij}\, d(\mathbf{Q}(\mathbf{a}_j,\, \mathbf{b}_i),\, \mathbf{x}_{ij})^2 \tag{1}$$

where

- $d(\mathbf{x}, \mathbf{x}')$ denotes the Euclidean distance between the image points represented by vectors $\mathbf{x}$ and $\mathbf{x}'$,
- $\mathbf{a}_j$ is the pose of camera at keyframe j and $\mathbf{b}_i$ the position of landmark i,
- $\mathbf{Q}(\mathbf{a}_j,\, \mathbf{b}_i)$ is the predicted projection of landmark i on the image associated to keyframe j,
- $\mathbf{x}_{ij}$ represents the observation of the i-th 3D landmark on the image of keyframe j and,
- v_{ij} stands for a binary variable that equals 1 if landmark i is visible in keyframe j and 0 otherwise.

Lets now consider that the map is divided into N submaps, each submap, say k, containing m^k keyframes and n^k landmarks, with $k = \{1, \ldots, N\}$. Then, (1) can be rewritten as

$$\min_{\mathbf{a}_j^l,\,\mathbf{b}_i^l} \sum_{k=1}^{N} \sum_{l=1}^{N} \left(\sum_{i=1}^{n^k} \sum_{j=1}^{m^l} v_{ij}^{kl}\, d(\mathbf{Q}(\mathbf{a}_j^l, \mathbf{b}_i^k),\, \mathbf{x}_{ij}^{kl})^2 \right) \tag{2}$$

where the combination of subscript i and superscript k refers to the i-th landmark of the k-th submap (e.g., $\mathbf{b}_i^k$), and similarly l over j refers to the j-th keyframe of the l-th submap (e.g., $\mathbf{a}_j^l$). Taking into account the observations shared between submaps, this expression can be written as

$$\min_{\mathbf{a}_j^l,\, \mathbf{b}_i^k} \sum_{k=1}^{N} \left(\underbrace{\sum_{\substack{l=1\\ l\neq k}}^{N} \sum_{i=1}^{n^k} \sum_{j=1}^{m^l} v_{ij}^{kl}\, d(\mathbf{Q}(\mathbf{a}_j^l,\, \mathbf{b}_i^k),\ \mathbf{x}_{ij}^{kl})^2}_{A} + \underbrace{\sum_{i=1}^{n^k} \sum_{j=1}^{m^k} v_{ij}^{kk}\, d(\mathbf{Q}(\mathbf{a}_j^k,\, \mathbf{b}_i^k),\ \mathbf{x}_{ij}^{kk})^2}_{B} \right) \tag{3}$$

where the term A stands for the reprojection error of those landmarks observed from keyframes of different submaps and the term B corresponds to the reprojection error of those landmarks observed form keyframes within the same submap. Both concepts are illustrated in figure 1.b. The first establishes the inter-connection between submaps which is represented by arcs connecting keyframes of different submaps (e.g. arc linking KF-2 and KF-11) and the second sets the intra-connection of the submap which includes the submaps inner arcs (e.g. arc linking KF-1 and KF-2).

If we are able to divide the map in such a way that the different submaps have few common observations, and assuming that the reprojection errors are independent of the map division, then A becomes negligible with respect to B. Thus, the global optimization can be approximated by

$$\sum_{k=1}^{N} \left(\min_{\mathbf{a}_j^k,\, \mathbf{b}_i^k} \sum_{i=1}^{n^k} \sum_{j=1}^{m^k} v_{ij}\, d(\mathbf{Q}(\mathbf{a}_j,\, \mathbf{b}_i),\ \mathbf{x}_{ij})^2 \right) \tag{4}$$

This approximation is equivalent to optimize each submap independently, which leads to a significant reduction of computational burden. In fact, this approximation is equivalent to the original expression (1) when there are no connections between submaps.

3.2 Map Partitioning Method

The approach proposed here to divide the map into coherent regions consists in grouping together those keyframes that observe the same features from the environment. For that, we consider the map as a graph whose nodes represent keyframes and the weight of the arcs are a measure of the common observations between them. There are two critical issues in this partitioning approach: first, the computation of the arc weights; and second, the criterion adopted to perform the partition itself. As for the first, the arc weights are assigned according to the Sensed-Space-Overlap (SSO), following our previous work [3], particularized for landmark observations. This simple but effective measure represents the information shared by two keyframes. It is calculated with the relation between the number of common landmark observations and the total number of landmarks

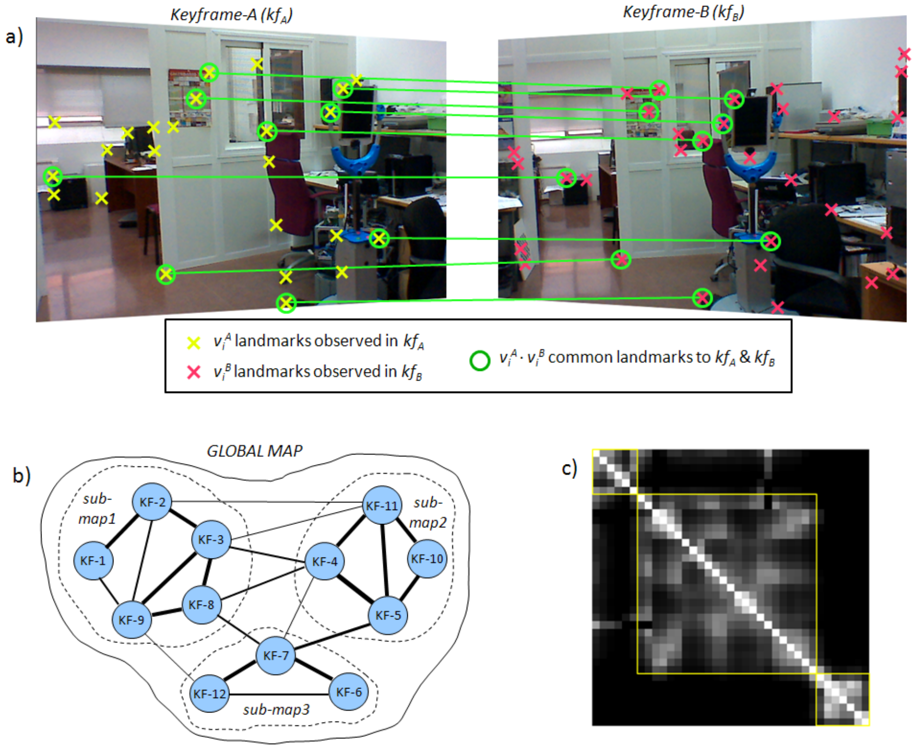

Fig. 1. a) Common observations between two keyframes. This is used to calculate the Sensed Space Overlap (SSO) (see equation 5). b) Graph-representation of the map where each node represents a keyframe and the arcs are weighed with the SSO calculated between keyframes (thicker arcs mean higher SSO). c) Example of SSO matrix, in which the brightness of the element ij represents the SSO between the keyframes i and j.

observed in both keyframes (see figure 1.a). This is expressed as

$$SSO\left(kf_A, kf_B\right) = \frac{\sum v_i^A \cdot v_i^B}{\sum v_i^A + \sum v_i^B - \sum v_i^A \cdot v_i^B} \tag{5}$$

where v_i^A and v_i^B, similarly to the definitions of the previous section, are binary variables that equal 1 if landmark i is observed in the keyframes kf_A and kf_B, respectively.

Regarding the criterion for partitioning the graph, we follow previous works [26, 2, 19] that apply the minimum normalized-cut (min-Ncut), originally introduced in [21]. The min-Ncut has the desirable property of generating balanced clusters of highly interconnected nodes, in our case clusters of keyframes that cover the same part of the environment. Figure 1 illustrates this concept: figure

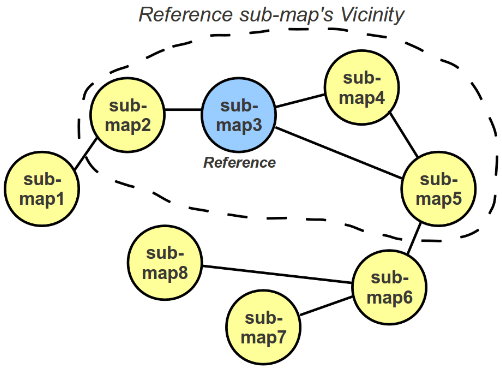

Fig. 2. Topological representation of the concept of submap vicinity.

1.a shows the common observations in a pair of keyframes whose arc weight is calculated with the SSO (see equation 5), and figure 1.b shows a map division into three submaps as produced by the min-Ncut procedure. Notice that the pairs of keyframes with higher SSO (thicker arcs) are grouped together. Figure 1.c shows the symmetrical SSO matrix corresponding to a different, larger map, where the keyframes are arranged according the min-Ncut to give rise to three groups of keyframes or submaps (matrix blocks).

It is important to notice that, in order to guarantee a scalable system when applying map partitioning to visual SLAM, the size of the submaps (i.e. number of keyframes) must be kept bounded. This requirement is not demonstrated mathematically here, but it is intuitive to see that as the camera explores new parts of the scene, the new keyframes will have low SSO values (if any) with distant ones in the map. Therefore, the min-NCut will produce new partitions when the system explores unobserved regions of the environment. This can be more clearly understood with the following example: lets consider the case where there are features that are always observed (e.g. the horizon when travelling by train, or when zooming in the scene, or traversing a corridor with the camera pointing in the movement direction) as the new keyframes are selected, they will introduce new features and therefore will reduce the minimum normalized-cut, resulting in the eventual partition of the map. The last two examples represent another advantage of our partition method, which produces natural multi-scale maps when the camera zooms. This insight is supported by all the experiments we have carried out during this work.

4 Dynamic division of PTAM's metric map

This section outlines the combination of our partition procedure and Parallel Tracking and Mapping (PTAM) [12]. PTAM is a monocular SLAM algorithm which performs online BA on keyframes, separating the tracking and mapping stages in two different threads to permit efficient real-time performance. This

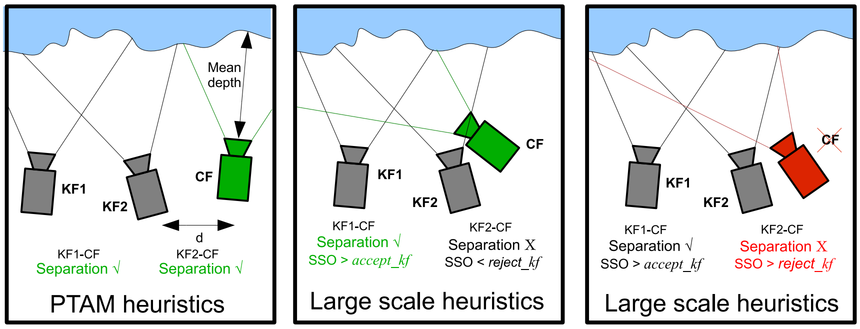

Fig. 3. Keyframe selection heuristics. a) PTAM's separation condition. b) and c) Keyframe acceptance and rejection heuristics, respectively, for large scale mapping. The thresholds used in our experiments for $accept\_kf$ and $reject\_kf$ are 0.2 and 0.7 respectively.

technique requires an initial map before it starts working automatically. Such initial map is acquired with a Structure from Motion procedure that involves user intervention to select two views with sufficient parallax. Once the initial map has been created, the system analyses the images retrieved by the camera to self-localize in the map, while the map is continuously optimized and augmented with new keyframes and landmarks. Such keyframes are selected according to some simple heuristics (see [12] for more details), and new landmarks are extracted through epipolar search between each new keyframe and its nearest keyframe in the map.

4.1 Keyframes selection in large environments

The keyframe selection criteria becomes an important aspect when PTAM is employed to build maps of large spaces. PTAM was designed for small environments (e.g. an office), where it works adequately with a hand-held camera which is waved sideways. PTAM employs a heuristic rule to select a new keyframe when there is a minimum separation between the current frame and the nearest keyframe in the map (i.e. Euclidean distance divided by the mean depth of the scene). This condition selects valid keyframes when the camera is moved sideways. But unlike in PTAM, we wish to explore big scenes and to construct large maps without being restricted to move sideways. Therefore, we have adapted this heuristic to select a keyframe when it provides useful information for mapping, by adding two more restrictions to the previous one for camera movement. Consequently, the current frame (CF) is selected as a new keyframe when:

- There exists a nearby keyframe which meets PTAM's separation condition with CF and which shares some information about the scene ($SSO > accept\_kf$).

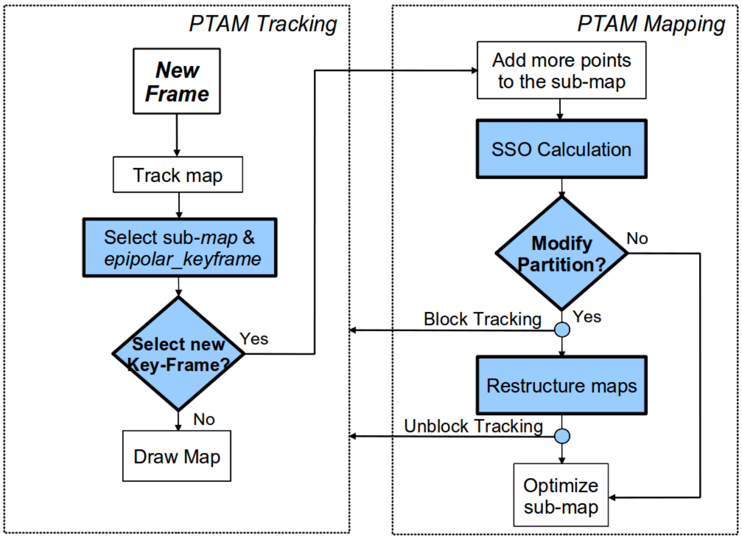

Fig. 4. Tracking and mapping threads of PTAM. Blue boxes correspond to the embedded stages to perform the map partitioning.

– There is not a nearby keyframe which does not meet PTAM's separation condition with CF and which shares much information about the scene ($SSO > reject\_kf$).

Figure 3 shows the adapted heuristics to select keyframes in large scale. PTAM's separation condition is shown in figure 3.a, where a keyframe is accepted when the Euclidean distance to the nearest keyframe divided by the mean depth of the scene is over some defined threshold. Figure 3.b shows the new acceptance condition, which selects the current frame if there exist at least one keyframe that fulfills PTAM's separation and whose $SSO > accept\_kf$ (KF1-CF). Figure 3.c shows the rejection condition, which rejects the current frame if there exist at least one keyframe that does not fulfil PTAM's separation and whose $SSO > reject\_kf$ (KF2-CF).

So, the acceptance condition prevents taking a new keyframe which shares little or no information with the map, while the rejection condition avoids selecting keyframes that are too similar to those already in the map. Hence, the combination of these two conditions permits selecting keyframes that provide new information to the map relaxing the movement constraints for nimble exploration of the scene.

4.2 Combination of Map Partitioning and PTAM

A scheme of the proposed partitioning method interacting with PTAM is depicted in figure 4. Our submapping procedure takes action in both of PTAM

Algorithm 1 Map Partitioning

M and KF are a submap and a keyframe respectively. SSO_M is the matrix containing the SSO values between all pairs of keyframes in the vicinity V. The *current_map* is the submap being tracked. *num_KF* is a keyframes counter and *N_part* is a parameter to control when the partition is to be reevaluated. A keyframe's *match_map* is the submap where it will be added, and a keyframe's *match_KF* is the keyframe used to find point correspondences.

After new keyframe *new_KF* is selected

```
num_KF++
Select match_map and match_KF
if match_map != current_map then
    num_KF = 0
end
Extract new map-points
Add a new row and a new column to SSO_M
for all submaps M_i ∈ V do
    for all keyframes KF_j of M_i do
        SSO_M ← SSO(new_KF, KF_j)
    end
end
if (num_KF % N_part) == 0 then
    Evaluate partition
    if partition is modified then
        Lock tracking thread
        for all submaps M_i ∈ V do
            Restructure M_i
        end
        Unlock tracking thread
        Update SSO_M
    end
end
```

threads. In the tracking thread, it selects the current submap and the nearest keyframe to the estimated pose after a new image is analyzed. In the mapping thread, after a new keyframe is selected and new landmarks are detected in it, the SSO is evaluated with respect to all the keyframes of the vicinity. Such vicinity includes all the submaps directly connected to the current submap (see figure 2).

The partitioning procedure comes into play after the SSO has been updated, then, the min-Ncut is evaluated, and if it results in a different partition, the map is rearranged. This procedure is described in algorithm 1. This partitioning method is applied dynamically as the map enlarges and may create new submaps as well as merge existing submaps to maintain coherency by grouping keyframes with high overlap. The result is a metric-topological map, where two different

Fig. 5. Experimental set up: laptop with attached camera.

topological areas will be connected by a rigid transformation if there are common observations between them.

The partitioning process, including SSO computation, min-NCut evaluation and map rearrangement depends on the number of keyframes and landmarks in the vicinity, taking up to 100 ms. in our experiments, which supposes a short time in comparison with the map optimization time.

4.3 Experimental Results

In this section we present some experiments which show the advantages, in terms of efficiency and scalability, of using the proposed metric-topological arrangement of the map instead of a single metric map. The experiments have been carried out using a Philips SPC640NC webcam, connected by USB to a linux-based laptop with an Intel Core2 Duo 2.4 GHz processor, 2Gb of memory and a nVidia GeForce-9400 graphics card. Figure 5 shows the set up of our monocular SLAM system.

A first experiment is aimed to illustrate the increase of efficiency in localization at frame rate. For that, we compare the time needed to project map points into the current frame with and without partitioning as the map grows. Both tests have been performed in the same environment, building maps composed of about 45000 points and 1000 keyframes, distributed in 52 submaps for the partitioning case. Figure 6 shows that the time with a unique map grows linearly with the number of map points, whereas with submapping, this time is bounded since only those points in submaps close to the camera are evaluated. This improvement in efficiency becomes more relevant when the map grows non-stop (note that this process is performed with each new frame captured by the camera).

The goal of a second experiment is to quantify the efficiency in the global optimization of the map with our submapping approximation. For that, we have run BA offline after every new keyframe is selected from a recorded video (that

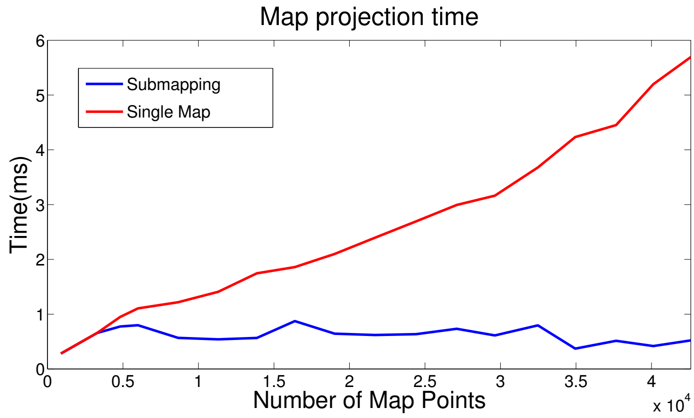

Fig. 6. Map projection time for localization with and without map partitioning.

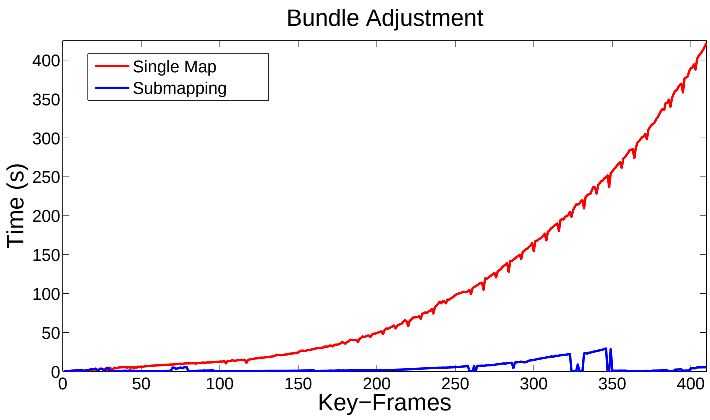

Fig. 7. Bundle adjustment computation time (offline) with and without partitioning.

is, sequential SfM), measuring the times of each BA completion with and without partitioning. At the end of these tests, the maps created were composed of about 22000 points and 400 keyframes, distributed in 9 submaps for the partitioning case. In order to compare both alternatives in the same conditions, we have included the time of partition management in the BA time for the partitioning test. Figure 7 shows the optimization times *vs.* the number of keyframes of the whole map for both cases.

As expected, for the case without partitioning, the computational cost follows an increasing polynomial trend with the number of keyframes. Conversely, when applying map partitioning, the computational burden is bounded since the BA is applied only on the current submap. For this case, we can observe some abrupt changes in the cost which are produced when the reference submap (the one where the system is localized) switches to a neighbor of different size. Figures

8.a and 8.b show the maps built with both alternatives (different colors represent different submaps in 8.b). We can verify visually their high similarity, and their good alignment, as a result of the continuous optimization previous to the map partition.

Additionally, we are interested in comparing the accuracy of the generated metric map. Due to the lack of a reliable metric to evaluate the maps quality, we have compared visually the different maps considering as ground truth the map obtained offline in the previous experiment (figure 8.a), which is the most accurate we can get. In the map obtained with PTAM (figure 8.c), we can appreciate some regions with depth errors and many outliers (e.g. landmarks detected behind physical walls). These inconsistencies are consequence of the premature interruption of global BA that happens when a new keyframe is selected, what leads to data association errors and the subsequent accuracy decrease with the map size. On the contrary, the map obtained with our approach (figure 8.d) presents no inconsistencies and considerably less outliers than the unique map solution (figure 8.c). This results from the higher efficiency of the submap local optimization, which optimizes regions with highly correlated observations to produce locally accurate submaps.

The results shown in this section have been supported in several tests performed under different conditions: exploring different rooms, re-visiting previous maps, traversing a corridor, zooming to get more detail of the scene, etc. The reader may refer to `http://youtu.be/-zK05EcOjX4` for a video that illustrates the operation of our submapping approach with PTAM in different environments.

5 Conclusions

This article presents an online submapping method which transforms a metric map into a metric-topological arrangement of it. This hybrid metric-topological structure improves the scalability of monocular SLAM in two aspects: first, the system rules out unnecessary metric information to perform more efficiently; second, it permits to use an approximation of BA to reduce computational cost while maintaining map consistency. Besides, the topological arrangement of the map is useful for other tasks, as loop closure, global localization or navigation. Experiments have demonstrated the potential of our approach to obtain efficient map representation in large environments. Future work will focus on exploiting the topological structure of the map for tasks as loop closure and relocalisation.

6 Acknoledgement

This work has been supported by the project “TAROTH: New developments toward a robot at home”, funded by the Spanish Government and the “European Regional Development Fund ERD” under contract DPI2011-25483.

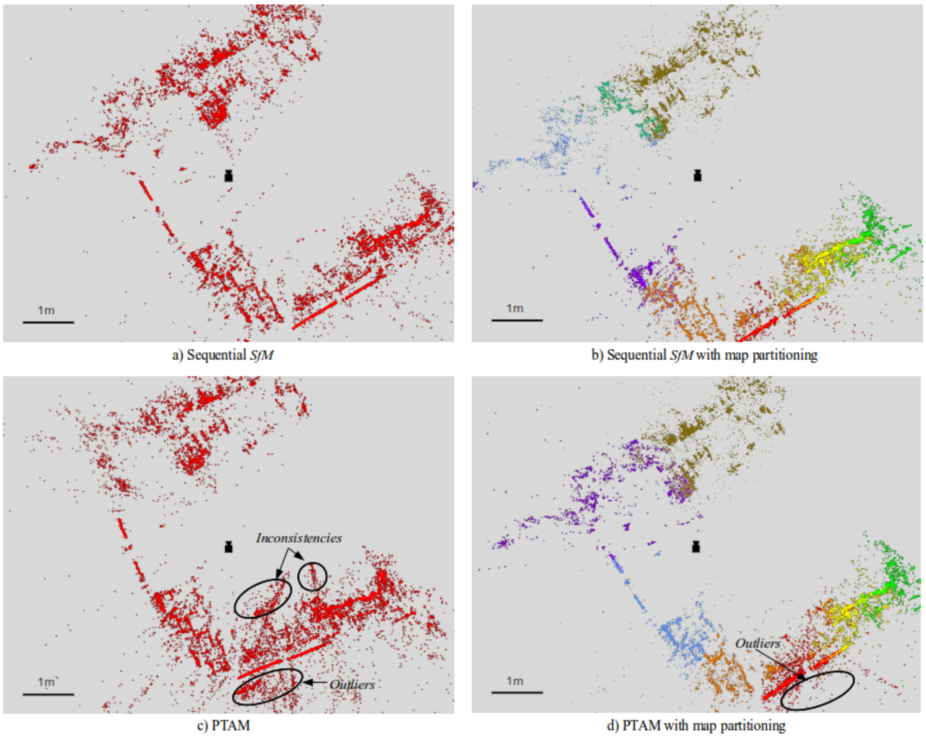

Fig. 8. Top view of maps generated in our experiments. All the maps are composed of more than 400 keyframes and 22.000 landmarks. The different colors in b) and d) represent different submaps.

References

1. Angeli, A., Doncieux, S., Meyer, J.A., Filliat, D.: Visual topological slam and global localization. IEEE International Conference on Robotics and Automation (2009)
2. Blanco, J.L., Fernández-Madrigal, J.A., González, J.: Towards a unified a bayesian approach to hybrid metric-topological slam. IEEE Transactions on Robotics **24**(2) (2008) 259–270
3. Blanco, J.L., González, J., Fernández-Madrigal, J.A.: Consistent observation grouping for generating metric-topological maps that improves robot localization. Proceedings of the IEEE International Conference on Robotics and Automation (2006) 818–823
4. Blanco, J., González-Jiménez, J., Fernández-Madrigal, J.: Sparser relative bundle adjustment (srba): constant-time maintenance and local optimization of arbitrarily large maps. In: IEEE International Conference on Robotics and Automation. (2013)
5. Davison, A.J.: Real-time simultaneous localisation and mapping with a single camera. Proceedings of the International Conference on Computer Vision (ICCV) (2003)

6. Eade, E., Drummond, T.: Monocular slam as a graph of coalesced observations. In: International Conference on Computer Vision. (2007)
7. Estrada, C., Neira, J., Tardos, J.: Hierarchical slam: Real-time accurate mapping of large environments. IEEE Transactions on Robotics **21**(4) (2005) 588–596
8. Fernández-Moral, E., González-Jiménez, J., Arévalo, V.: Creating metric-topological maps for large-scale monocular slam. In: International Conference on Informatics in Control, Automation and Robotics (ICINCO). (2013)
9. Fernández-Moral, E., Mayol-Cuevas, W., Arévalo, V., González-Jiménez, J.: Fast place recognition with plane-based maps. In: IEEE International Conference on Robotics and Automation. (2013)
10. Galindo, C., Saffiotti, A., Coradeschi, S., Buschka, P., Fernández-Madrigal, J., González, J.: Multi-hierarchical semantic maps for mobile robotics. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2005) 2278–2283
11. Holmes, S., Sibley, G., Klein, G., Murray, D.: A relative frame representation for fixed-time bundle adjustment in monocular sfm. In: IEEE International Conference on Robotics and Automation. (2009)
12. Klein, G., Murray, D.W.: Parallel tracking and mapping for small ar workspaces. In Proceedings of the International Symposium on Mixed and Augmented Reality (2007)
13. Konolige, K.: Sparse sparse bundle adjustment. In: British Machine Vision Conference. (2010)
14. Lim, J., Pollefeys, M., Frahm, J.M.: Online environment mapping. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2011)
15. Newman, P., Leonard, J., Soika, M., Feiten, W., Teller, S.: An atlas framework for scalable mapping. In: IEEE International Conference on Robotics and Automation. Volume 2. (2003) 1899–1906
16. Ni, K., Dellaert, F.: Multi-level submap based slam using nested dissection. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. (2010)
17. Ni, K., Steedly, D., Dellaert, F.: Tectonic sam: Exact, out-of-core, submap-based slam. In: IEEE International Conference on Robotics and Automation. (2007)
18. Nistér, D., Naroditsky, O., Bergen, J.R.: Visual odometry. Proc. IEEE International Conference on Computer Vision and Pattern Recognition (2004) 652–659
19. Rogers, J.G., Christensen, H.I.: Normalized graph cuts for visual slam. IEEE/RSJ International Conference on Intelligent Robots and Systems (2009)
20. Savelli, F., Kuipers, B.: Loop-closing and planarity in topological map-building. In: IEEE/RSJ International Conference on Intelligent Robots and Systems. Volume 2. (2004) 1511–1517
21. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(8) (2000) 888–905
22. Sibley, G., Mei, C., Reid, I., , Newman, P.: Adaptive relative bundle adjustment. In Robotics Science and Systems Conference (2009)
23. Strasdat, H., Davison, A., Montiel, J., , Konolige, K.: Double window optimisation for constant time visual slam. IEEE International Conference on Computer Vision (ICCV) (2011)
24. Strasdat, H., Montiel, J.M.M., Davison, A.J.: Scale drift-aware large scale monocular slam. Robotics: Science and Systems (2010)
25. Thrun, S.: Learning metric-topological maps for indoor mobile robot navigation. Artificial Intelligence **99**(1) (1998) 21–71

26. Zivkovic, Z., Bakker, B., Krose, B.: Hierarchical map building using visual landmarks and geometric constraints. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2005) 2480–2485