Reviewer 2

Dear sir, Thanks for your time and effort in reviewing the paper. Below is a detailed response to all your suggestions. The changes in the paper have been set in blue font to ease their detection.

The purpose of this paper is to achieve a robust and fast visual localization in environments where artificial markers have been placed. The paper has practical applications and sufficient experiments However, the method of simply eliminating key points is insufficient in terms of innovation. Consequently, I have some suggestions for the authors.

1. In the part of the “System overview”, You can add a definition of the coordinate system.

As suggested, the following text has been added (line 192-193):

“Note that a reference system describes objects' relative positions (location and orientation) in a specific space.”

2. The descriptions of existing terms or definitions can be simplified.

It is unclear which terms or definitions the reviewer refers to, so we are unsure what changes to make to the paper. In any case, our definitions aim at being easy to understand by any reader, so we consider keeping them as they are not harmful.

3. In line 410, you mentioned that the data used in the experiment were collected on your university campus. However, in line426, Is the description of “have employed the SPM dataset”? Where exactly does the first 6 indoor datasets come from?

All datasets have been recorded at our University Campus. While the corridors and outdoor areas have been recorded for this work, indoor areas (corridors and ceiling) correspond to a laboratory area already created in our previous work [6].

For clarity, the text has been modified (line 408-412):

“For the experiments, 22 video sequences were used from different zones. The camera captured the sequences while moving between 4 and 6 km/h in each environment. The first 6 sequences belonged to the SPM dataset \cite{SPMSLAM} and were recorded indoors, spotting at walls and ceilings. The remaining 16 sequences include 2 from the corridors and lobby of a building and 14 from various outdoor locations on our University Campus, see Table “

4. I noticed that in your experiments, you only tested the algorithm within the same environment after creating a base map, limiting its potential applications. Could you conduct experiments where the algorithm is applied to overlapping but distinct environments after the base map has been created?

The reviewer is correct. Our work mainly focuses on creating maps of areas where an autonomous agent (e.g., a robot) will eventually move to do a task. This work is not focused on exploring unknown areas while doing the tracking task since it is dangerous in many use cases. Nevertheless, our recordings include different viewpoints on the same area, thus exploring different paths on the map.

As future work, we find it interesting to expand the *base map* to incorporate previously unexplored areas.

This sentence has been added to the text (line 572-574):

“Finally, the experiments have been conducted in environments where the base map was created. As future work, it would be desirable to extend this functionality, allowing the base maps to be extended to new areas not explored in the first phase.”