

# Wearable Omnidirectional Vision System for Personal Localization and Guidance\*

A. C. Murillo, D. Gutiérrez-Gómez, A. Rituerto, L. Puig, J. J. Guerrero  
DIIS - Instituto de Investigación en Aragón. Universidad de Zaragoza, Spain.  
{acm, danielgg, arituerto, lpuig, jguerrer}@unizar.es

## Abstract

*Autonomous navigation and recognition of the environment are fundamental abilities for people extensively studied in computer vision and robotics fields. Expansion of low cost wearable sensing provides interesting opportunities for assistance systems that augment people navigation and recognition capabilities. This work presents our wearable omnidirectional vision system and a novel two-phase localization approach running on it. It runs state-of-the-art real time visual odometry adapted to catadioptric images augmented with topological-semantic information. The presented approach benefits from using wearable sensors to improve visual odometry results with true scaled solution. The wide field of view of catadioptric vision system used makes features last longer in the field of view and allows more compact location representation which facilitates topological place recognition. Experiments in this paper show promising ego-localization results in realistic settings, providing good true scaled visual odometry estimation and recognition of indoor regions.*

## 1. Introduction

The ability to navigate effectively in the environment and recognize places, objects and signals is fundamental and natural for people. However the solution is not always obvious and turns more complicated when we are in unknown or difficult to access environments or in daily situations for visually impaired people. In the last decades plenty of research in the robotics and computer vision fields has tried to replicate these two important tasks in realistic scenarios. At the same time, quickly evolving technologies are providing us with a wide range of high performance miniaturized computers and sensors. These conditions have raised opportunities for many new assistance applications for people through the use of wearable intelligent sensors. Broad range

\*This work has been supported by the project DPI2009-14664-C02-01, the I3A Fellowship Program and FEDER.

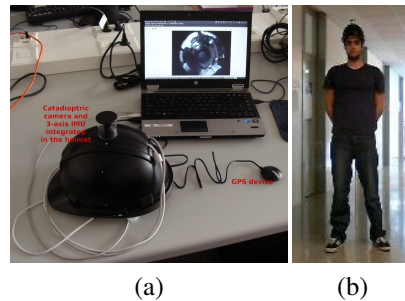


Figure 1. (a) Wearable omnidirectional vision system built on a helmet. (b) User wearing the helmet during experiments.

of different platforms and sensors have been used but probably cameras provide one of the richest data and lower cost platforms. Since initial prototypes such as [13], we find plenty of proposed wearable sensor applications [15]. Recently we find wearable systems proposed for recognizing daily activities [27, 26]; other recent works, closer to ours, are oriented to specific assistance such as visually impaired or elderly recognition and navigation aid systems [10, 14].

**Contributions.** This paper presents our wearable omnidirectional vision system (see Fig. 1), the online available datasets<sup>1</sup> acquired with it and a method for personal localization, which establishes the basic framework for the person guidance. Visual odometry is running in real time, and topological localization is efficiently obtained every few meters. Two steps are being developed to get fully functional system: connecting both modules in real time and augmenting path and localization provided on the topological map with voice commands for the user.

Our prototype exploits the advantages of omnidirectional catadioptric systems, a novelty compared to related works, which allows more compact and robust representation of the environment. The proposed localization system builds on a recently proposed real time omnidirectional SLAM method [8]. In this paper we validate its suitability for indoor environments, where catadioptric vision systems suffer bigger distortions, and enhance it with topological-semantic information required for human guidance.

<sup>1</sup><http://robots.unizar.es/omnicam>

**Human-centered localization vs. robot visual localization.** Our method adapts typical hierarchical localization approaches from mobile robotics to human-centered applications. We run an accurate metric ego-motion estimation, transparent to the user, and a topological/semantic localization, necessary to guide user motion and provide environment information. The presented approach takes into account several particularities when camera is worn by a person as opposed to mounted on an robotic platform. On one hand, a person may not need accurate metric estimation of its current position, but more general/semantic information of the region traversed. On the other hand, it is still relevant to keep accurate track of the person motion through the visual odometry, to provide necessary warnings, information and commands at the right location. Finally, prior knowledge on people motion provides interesting hints to accurately estimate the scale of the visual odometry. This scale is typically unknown without prior information and solutions are usually given up to a scale factor. Here this is solved thanks to the fact of using a wearable platform.

**Vision system for personal localization approach.** Fig. 2 summarizes our approach. It consists of two modules. The first one executes a six degrees of freedom real-time SLAM algorithm, which estimates the odometry of the person while walking and allows us to keep track of the user trajectory (Fig. 2b). This module provides a true scaled and accurate metric estimation and sends a request to the topological module (Fig. 2a) every two meters walked. The second module runs a topological localization algorithm which can provide the user with navigation instructions, i.e., path across the topological graph to reach given goal, and information about the area of the environment that is being traversed. This module requires a visual memory or model of the environment (Fig. 2c), which is previously acquired with the same vision system and includes labels and topological information of the regions in the environment.

## 2. Related Work

There is an increasing interest in devise robotics systems for personal assistance, and as already mentioned, current miniaturization of sensors has also raised the amount of wearable assistance systems. We find earlier systems that adapt robot localization methods to wearable omnidirectional vision systems [22]. Many other general vision-based approaches have been adapted to human navigation assistance, usually considering cameras enhanced with additional sensors to measure sound [10] or depth [14]. Our method adapts the use of omnidirectional images as a key characteristic. Many localization systems take advantage of wide field of view cameras to acquire more compact visual models and perform robust operations with just one wide field of view image, e.g., in [7] an approach for topological

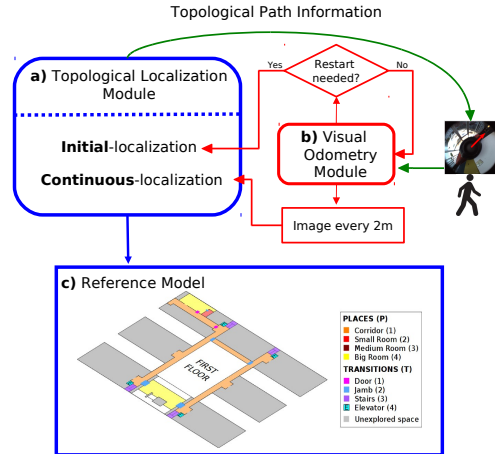


Figure 2. Diagram of the localization and guidance approach.

mapping and navigation using a catadioptric vision system is presented. Another related approach is the vision based navigation assistant presented in [11], which builds a topological map using a wide field of view system composed of four cameras mounted on the user’s shoulder.

A natural approach to vision-based navigation is to explicitly estimate metrical ego-motion from images. The problem of recovering both, the camera motion and the structure of the environment is known as Structure from Motion in computer vision literature and Simultaneous Localization and Mapping (SLAM) in robotics literature. Achieving efficient performance remains as a long-standing problem because even the best metric SLAM systems degrades with time. In order to cope with the complexity and scale of real environments, topological approaches [7, 3] have been proposed. This type of approach must build an annotated topological map of the environment by capturing location dependent scene appearance. Moreover, topological annotations lead to further semantic environment information that facilitates human interaction.

**Topological and Semantic Visual Localization.** Topological modeling of the environment has been studied for a long time. Recently, interest has increased due to its advantages to deal with large environments and to augment the models with semantic information. Different approaches are used to augment topological models with additional information, e.g., with human supervision to achieve representations closer to human concepts [30], or with weaker human supervision to learn semantic models from a few initial labeled samples [23].

Topological localization consists of finding the most similar reference image in the model with regard to our current view, in order to retrieve the stored/learned information related to that location. Vision based topological localization is closely related to the more general problem of place or scene recognition, such as this recent work for rec-

ognizing types of indoor locations [20]. Topological maps are usually built on top of a hierarchy of different map levels [12], e.g., a global topological map that connects smaller local metric maps [29].

**Omnidirectional Visual SLAM.** Visual SLAM has been addressed both using monocular and stereo systems. Stereo systems allow the map and visual odometry of the camera to be completely estimated through triangulation of the observed landmarks of the scene. On the other hand monocular systems only allow an up-to-scale scene reconstruction due to the impossibility of directly measuring landmarks depth from just one image. Nevertheless, Monocular Visual SLAM [4] presents high potential, since these cameras are cheaper, compact and easier to calibrate.

An improvement to monocular visual SLAM is to increase the field of view of the system by using omnidirectional cameras. Features typically last longer in the field of view, specially when large camera rotations occur. The increased lifespan of the features on the image translates in a better estimation of the position of the features on the map, a lower need to initialize new features and an increased robustness. However omnidirectional cameras, in particular catadioptric systems, involve a more complex projection model, important image deformation, distortion and variable scale in the image. Some visual SLAM approaches using omnidirectional cameras have been proposed [28, 25, 21]. All these works show advantages of omnidirectional cameras to deal with this problem.

### 3. Topological/Semantic Localization Module

The topological localization module from our system relies on a previously built model. First initial global localization is computed, and the user can ask the system for a path to reach any other region in the modeled environment, obtained following standard graph search strategies. The topological model could be automatically generated, but in our current version is based only on human supervision.

Given current view, topological localization step consists of finding the most similar image among the reference views in the environment model. Key issues are how to represent and compare the similarity between images, cf. section 3.1, and how to select which images from the reference set are possible localization candidates, cf. section 3.2.

#### 3.1. Image representation

For more efficient and compact representation, we use the Gist based descriptor for catadioptric images, detailed in [16]. Due to the camera orientation in the helmet, if we obtain four partitions as shown in Fig. 3, they correspond to the main motion directions, according to the Manhattan World Assumption. With this representation, the omnidirectional image Gist  $\mathbf{g}$  is composed by four conven-

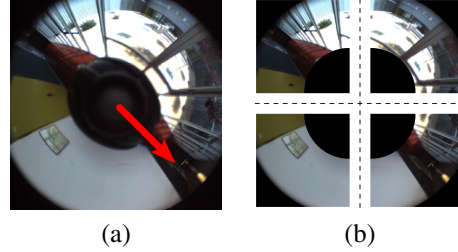


Figure 3. (a) Omnidirectional image acquired with the Wearable OmniCam system. Red arrow points front direction of the helmet. (b) Partitions performed to obtain the Gist based description.

tional Gist descriptors [18], one computed for each image part (front, left, back and right):  $\mathbf{g} = [g_f, g_l, g_b, g_r]$ . The similarity between two images using this representation is obtained based on the Euclidean distance between the descriptors. We compute the minimum distance that can be obtained between one image and the four possible permutations of the four partitions of the second image. These permutations correspond with the four possible alignments of the sectors of the image and will provide us with the best alignment of the two evaluated images. Being  $\mathbf{g}$  and  $\mathbf{g}'$  the descriptors of two images, the distance between them is:

$$dist(\mathbf{g}, \mathbf{g}') = \min_m (d_e(\mathbf{g}, \pi_m(\mathbf{g}'_{flbr}))), \quad (1)$$

where  $\pi_m(\mathbf{g}'_{flbr})$  is the  $m^{th}$  circular permutation of the descriptor  $\mathbf{g}'$  component vectors ( $m = 1, 2, 3, 4$ ) and  $d_e$  the Euclidean distance between the Gist descriptors of two omnidirectional images.

For more accurate computations we use local features, in particular standard extraction and correspondence search of SURF [1], which has been shown to provide good performance on the same kind of catadioptric images we use [17].

#### 3.2. Topological Localization and Guidance

Topological localization of the user can be obtained on two different situations: either all reference images should be considered as possible localization candidates (global localization) or prior knowledge allows a reduced set of candidates (continuous localization).

**Global Localization.** When the system is initialized or re-started we do not consider any prior about current localization. Re-start can occur when the visual odometry module is inconsistent or when it is lost. The user could be anywhere and current view is compared to all reference images through a hierarchical process:

1. A small set of top- $k$  candidates similar to the query is found using representation and similarity from (1).
2. The  $k$  selected candidates are re-ranked according to the amount of local SURF feature correspondences (obtained with standard nearest neighbor ap-

proach). Reference candidate image with more correspondences is given as the topological location.

**Continuous Localization.** While the user is moving, we assume that between two consecutive queries the user can stay in the same topological region or move to one of the neighboring regions (connected by at most two steps) in the topology. Therefore, we have a reduced set of candidate locations. Similarity from query to these candidates is evaluated only using Gist descriptors as defined in (1).

Continuous topological information would allow the system to check if current location is on the path given to the user to achieve the goal or if the system should provide some reminder, new command or re-planification.

## 4. Visual Odometry Module

As described, our proposed SLAM module intends to keep track of the motion performed by the person to be able to guide it through the environment. This section describes how we compute the visual odometry of the person, since we do not keep track of the estimated landmarks locations (map) but only the camera (person) location at each time.

We use the approach presented in [8], that proposes to combine state of the art robust EKF monocular SLAM [2] with an omnidirectional sensor. As catadioptric images involve a high deformation of the world, linear transformations are applied to the features to improve the matching results. This transformation consists on a rotation and scaling of the feature patch. The rotation is computed from the variation of the polar angle of the feature in the image, and the scale patch factor is computed as follows:

$$k = \frac{D_{init}}{D_{current}} \frac{\xi + f(\xi, \frac{R_{im,init}}{\gamma})}{\xi + f(\xi, \frac{R_{im,current}}{\gamma})}, \quad (2)$$

where  $D_t$  is the distance from the landmark to the camera location at moment  $t$ ,  $R_{im,t}$  is the distance of the projected feature to the principal point of the image and  $\xi$  and  $\gamma$  are camera-mirror parameters. This formula encodes not only the natural change of the patch size due to the variation of the distance between the corresponding landmark and the camera but also the change of the scale in the radial direction of the image induced by the mirror.

**True scaled visual odometry.** It was previously stated that one of the problems of monocular SLAM is the impossibility to observe the scale of the scene. This problem leads to a scale drift along the camera trajectory which result in a deformation of the final map and trajectory reconstruction. Some authors have proposed different ways to solve the scale problem of monocular SLAM either by combining the visual data with information from other sensors [6] or using some kind of prior distance [24].

To obtain a true scaled estimate of the visual odometry we apply the algorithm presented and detailed in [9]. This approach takes advantage of the induced head vertical oscillation during walking, whose frequency matches up with the step frequency. This is an interesting approach only valid for wearable cameras.

## 5. Experiments

This section presents the experiments run to validate our localization modules in a typical indoor environment and provides more detailed description of the prototype designed and dataset used.

### 5.1. Wearable omnidirectional camera

The wearable omnidirectional vision system presented consists of a set of small and light sensors mounted on a helmet that will be carried by a person. The whole prototype (Fig. 1) consists of a laptop, to be carried on a backpack, that runs all the processing and controls the following set of USB sensors, which do not require any extra power supply:

- An omnidirectional catadioptric camera. The main sensor of the system is an omnidirectional camera manufactured by Vstone<sup>2</sup>, with dimensions 47mm×80mm×47mm. This catadioptric camera provides images with resolution of 1024x768 at a frame rate of 15fps.
- An inertial measurements unit (IMU). This IMU manufactured by Xsens<sup>3</sup> contains 3 accelerometers, 3 gyroscopes and 3 magnetometers. It is attached to the catadioptric camera and provides the attitude of the camera.
- A GPS. When operating outdoors, this sensor can provide the position of the system with up to 5m error, however we can not make proper use of this sensor in the indoor experiments presented next in this paper.

### 5.2. Dataset and experimental settings

This dataset has been acquired in a three floor building and traverses different types of regions: corridors, research laboratories, offices, classes, etc. Acquisitions have been performed by a person wearing the helmet, so the dataset suffers the typical motion of a person walking. Sequences can be downloaded from the web, as well as additional data acquired with the prototype. For this work the only sensor used is the calibrated omnidirectional camera [19].

As described in Sec. 3, the topological/semantic localization module needs a visual model of the environment. The construction of this model is performed with a separated *training* acquisition obtained more than six months before the testing sequences (*may2011\_sequence*). Figure 4 shows the map of the three floors of the building highlighted with different colors, depending on the type of area traversed

<sup>2</sup><http://www.vstone.co.jp>

<sup>3</sup><http://www.xsens.com/>

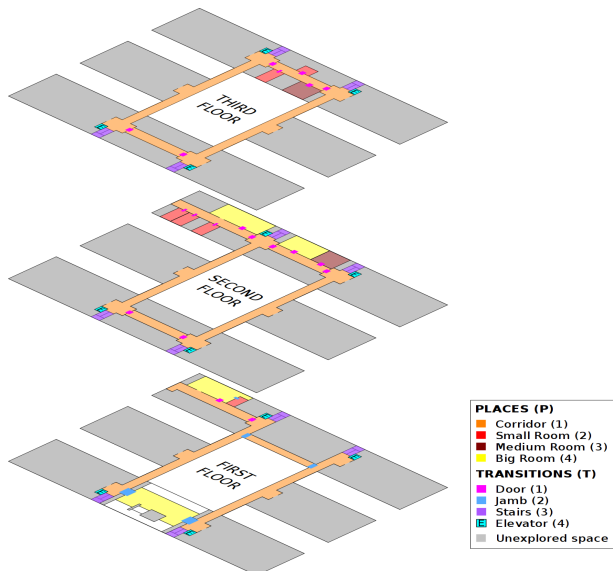


Figure 4. Training data. Map of the building where the dataset has been acquired. Different colors show different area type traversed.

during this training sequence. The gray areas are parts not covered by the training data. This sequence consists of 20905 omnidirectional images with a manual ground truth labeling of the type of areas traversed. We consider the main spaces of a building, like corridors or rooms, as *Places*. *Transitions* label comprises all the areas joining different *Places*: doors, jambs, stairs and elevators. The areas labeled as *Transitions* start and end about 0.5 meters before and after the *Transition* has been crossed.

*Test* images come from an additional acquisition (*march2012\_sequence*) during a different trajectory consisting of 7027 frames. Ground truth for the topological/semantic localization is manually assigned similarly to the training data labels. Ground truth for the SLAM module evaluation was obtained from the plans of the building and Google Maps tools<sup>4</sup>. The test sequence is divided in two parts (Fig. 5) whose lengths are 392 m and 75 m. Division is made at the point where the SLAM module shows an excessive growth of the uncertainty ellipses of the features on the image or an obvious deviation from the true path. Then, the system needs to perform another global localization and start a new SLAM estimation.

Note that we do not get a global metric and accurate localization for all parts, since different SLAM results are never merged. However, it is not needed, because we only need to keep track of the topological regions traversed to guide the person with semantic commands. Therefore, main interest in the SLAM is to achieve good local estimation to keep track of current user movements and scale to provide equally separated images to the topological localization.

<sup>4</sup><http://support.google.com/maps/>

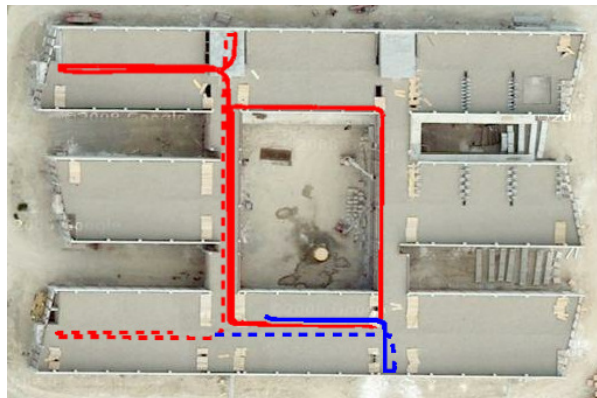


Figure 5. Test data. Ground truth corresponding to the first part (red) and the second part (blue) of test trajectory (solid line indicates 1st floor, dashed line indicates 2nd floor).

### 5.3. Performance evaluation

**Visual odometry module.** An important goal of our experiments was to achieve good visual odometry estimation to avoid using frequently the costly global localization (global localization needs to perform much more and slower comparisons than continuous localization) and to be able to track the user movements in the environment. Visual odometry is obtained with the scaling algorithm detailed in [9]. Following this method, to estimate the walking speed as a function of the step frequency, we need a prior computation of parameters  $\alpha$  and  $\beta$  for the user wearing the camera. We measured the time spent to walk a fixed distance at different step frequencies and performed a power fitting on the experimental data to obtain  $\alpha = 0.329$  and  $\beta = 1.534$ .

The raw un-scaled SLAM estimation used our modification for catadioptric systems of the real-time approach for conventional cameras [5]. The trajectory is computed in real time, each step of the SLAM algorithm takes an average of 0.058s, while camera acquires frames every 0.067s.

Fig. 6 shows the final reconstruction of the visual odometry without using topological information and applying our scaling algorithm. Although not fitting completely with the ground truth, mainly due to the accumulation of errors in the estimation of rotations, it can be observed that the final visual odometry estimation preserves the shape and size of the ground truth trajectory. Notice how the scale is quite accurately estimated, especially with respect to the raw estimation from the SLAM algorithm. Also the comparison with the estimation resulting from applying a uniform scale factor shows that the method to estimate the scale is able to cope with a large scale drift along the trajectory. Quantitative measure of these improvements is given by the estimated total trajectory length: our approach estimates 434m and 95m for part 1 and 2 respectively, very close to those 392m and 75m given by measurement tools used as ground truth: respectively. It is a significant improvement

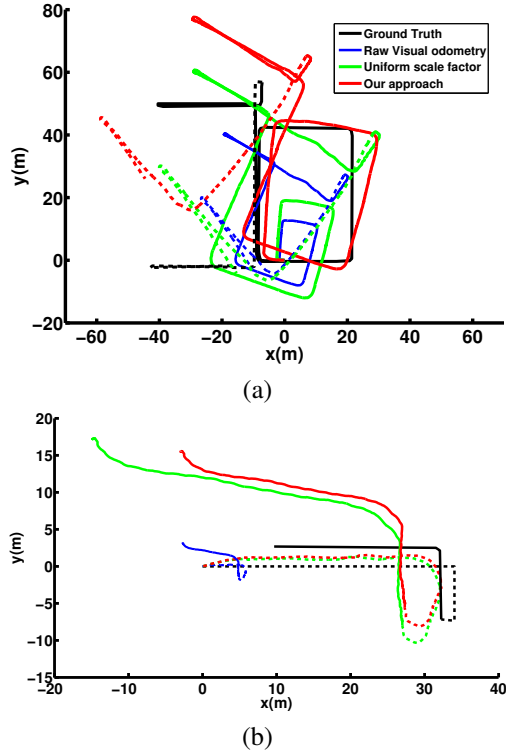


Figure 6. Visual odometry estimation from different approaches. (a) part 1. (b) part 2.

compared to the raw odometry estimations, with lengths of 275m and 20m respectively. Besides, note that a person trajectory is barely a straight line, but this is not considered by the measurement tools, giving usually shorter measurements.

**Semantic/Topological localization.** This section summarizes the topologic-semantic localization evaluation. Quantitative evaluation is obtained by measuring how many test images got assigned the correct topological cluster.

First we evaluate the *global localization* algorithm at each test image, i.e., we compare every test image with all reference images for different values of  $k$ , as described in Section 3.2. We observe that  $k = 10$  provides the best results, 88% and 42% correct localization tests for part 1 and part 2, respectively. Larger or smaller values of  $k$  provided lower performance, then  $k = 10$  will be used in following experiments. We should note that the global localization step is required only at the beginning of the processing of each part, and it provides the correct topological location in the two starting points on our dataset as shown in Fig. 7 (green arrows point to the starting location for each part).

*Continuous localization* is also evaluated at every test image. Table 1 shows the results using different amount of reference images. The query image is compared only to every *step* images from the reference trajectory.

Fig. 7 shows the topological localization obtained apply-

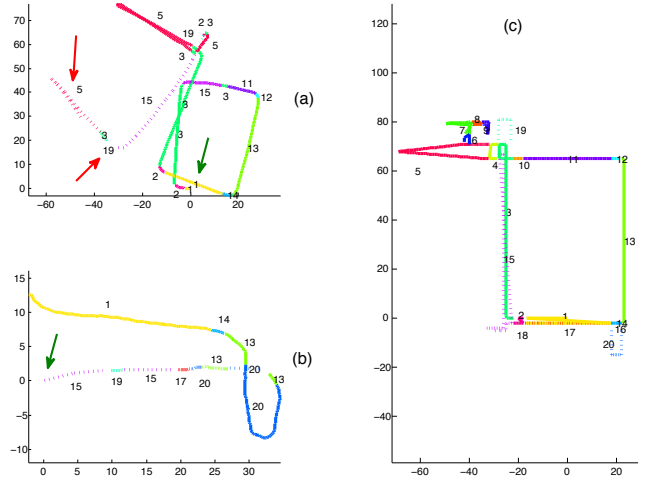


Figure 7. Topological localization obtained for the two test parts (a) and (b) and the ground truth available (c) with the reference topological regions. Each color and number points the reference topological region assigned to the nearby group. As previously, dashed lines show locations in the second floor.

	Part 1	Part 2
$step = 5$	87.9%	64.6%
$step = 10$	79.1%	25%

Table 1. Correct continuous localization for the 2 test parts ( $k=10$ ).

ing the whole approach. Global localization is run at the beginning of each part, and continuous localization is run for every frame that visual odometry module requests while the user moves. We consider each set of contiguous images which obtained the same topological label as a cluster. Each number plotted in the figure corresponds to the reference cluster assigned to each of these groups. By comparing with the ground truth diagram (see Fig. 7c), we can observe a few errors in the topological classification. In part 1 (Fig. 7a), there are eight misclassified clusters, but these errors correspond to short clusters, so confusion does not last long. Red arrows show important points where errors occur. First error occurs at a corridor that was not explored in reference trajectory, it gets misclassified as cluster 5, and 5 actually corresponds to other similar corridor. The other red arrow points a case where we misclassify as cluster 19 (stairs), but it actually corresponds to other nearby stairs. In part 2 (Fig. 7b), we can also observe that incorrect group classification do not last too long. The only long error occurs at the beginning, where the system takes long to realize that it has left the initial cluster.

## 6. Conclusions and Future Work

We have described a wearable catadioptric vision system designed to research on personal assistance applications. This work presents our novel proposal for people ego-localization and initial guidance steps, based on state

of the art methods from autonomous mobile robots. It proposes how to adapt and improve them thanks to the advantages of wearable sensors and omnidirectional vision. Our initial experiments in this paper show promising ego-localization results obtained in realistic settings. We also provide a scaled visual odometry estimation, that helps the system to track the movements of the user and the recognition of different indoor regions. This is necessary for more intuitive interaction with users and their guidance around the environment. Future steps require real-time integration of all modules with additional human-machine interaction modules in a real time system.

## References

- [1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Comput. Vis. Image Underst.*, 110:346–359, June 2008. **3**
- [2] 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*, 27(5):609–631, October 2010. **4**
- [3] M. Cummins and P. Newman. Highly scalable appearance-only SLAM - FAB-MAP 2.0. In *Proc. of Robotics: Science and Systems (RSS)*, 2009. **2**
- [4] A. J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Int. Conf. on Computer Vision*, 2003. **3**
- [5] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:1052–1067, June 2007. **5**
- [6] A. Eudes, M. Lhuillier, S. Naudet-Collette, and M. Dhome. Fast odometry integration in local bundle adjustment-based visual slam. *ICPR*, pages 290–293, 2010. **4**
- [7] T. Goedemé, M. Nuttin, T. Tuytelaars, and L. Van Gool. Omnidirectional vision based topological navigation. *International Journal of Computer Vision*, 74(3):219–236, 2007. **2**
- [8] D. Gutierrez, A. Rituerto, J. M. M. Montiel, and J. J. Guerrero. Adapting a real-time monocular visual slam from conventional to omnidirectional cameras. In *11th OMNIVIS, held with ICCV*, 2011. **1, 4**
- [9] D. Gutiérrez-Gómez, L. Puig, and J. Guerrero. Full scaled 3D visual odometry from a single wearable omnidirectional camera. Technical report, University of Zaragoza, 2012. **4, 5**
- [10] M. Havlena, A. Ess, W. Moreau, A. Torii, M. Jancosek, T. Pajdla, and L. Van Gool. Awear 2.0 system: Omnidirectional audio-visual data acquisition and processing. In *CVPR Workshops*, pages 49–56, June 2009. **1, 2**
- [11] O. Koch and S. Teller. Body-relative navigation guidance using uncalibrated cameras. In *Int. Conf. on Computer Vision*, pages 1242–1249, 2009. **2**
- [12] B. Kuipers. The Spatial Semantic Hierarchy. *Artificial Intelligence*, 119(1-2):191–233, 2000. **3**
- [13] S. Mann. Wearable computing: A first step toward personal imaging. *Computer*, 30:25–32, 1997. **1**
- [14] S. Mann, J. Huang, R. Janzen, R. Lo, V. Rampsad, A. Chen, and T. Doha. Blind navigation with a wearable range camera and vibrotactile helmet. In *ACM Int. Conf. on Multimedia*, pages 1325–1328. ACM, 2011. **1, 2**
- [15] W. Mayol-Cuevas, B. J. Tordoff, and D. W. Murray. On the choice and placement of wearable vision sensors. *IEEE Trans. on Systems Man And Cybernetics*, 39(2):414–425, March 2009. **1**
- [16] A. C. Murillo, P. Campos, J. Kosecka, and J. J. Guerrero. Gist vocabularies in omnidirectional images for appearance based mapping and localization. In *10th OMNIVIS, held with Robotics: Science and Systems (RSS)*, 2010. **3**
- [17] A. C. Murillo, J. J. Guerrero, and C. Sagüés. Surf features for efficient robot localization with omnidirectional images. In *IEEE/RSJ Int. Conf. on Robotics and Automation*, pages 3901–3907, 2007. **3**
- [18] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001. **3**
- [19] L. Puig, Y. Bastanlar, P. Sturm, J. J. Guerrero, and J. Barreto. Calibration of central catadioptric cameras using a dlt-like approach. *International Journal of Computer Vision*, 93(1):101–114, March 2011. **4**
- [20] A. Quattoni and A. Torralba. Recognizing indoor scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 413–420, June 2009. **3**
- [21] A. Rituerto, L. Puig, and J. J. Guerrero. Visual slam with an omnidirectional camera. In *20th International Conference on Pattern Recognition*, pages 348–351, 2010. **3**
- [22] W. Rungtanyotin and T. E. Starner. Finding location using omnidirectional video on a wearable computing platform. In *Int. Symp. on Wearable Computing*, pages 61–68, 2000. **2**
- [23] O. Saurer, F. Fraundorfer, and M. Pollefeys. Visual localization using global visual features and vanishing points. In *Conference on Multilingual and Multimodal Information Access Evaluation (CLEF 2010)*, 2010. **2**
- [24] D. Scaramuzza, F. Fraundorfer, M. Pollefeys, and R. Siegwart. Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints. In *Int. Conference on Computer Vision*, 2009. **4**
- [25] D. Scaramuzza, F. Fraundorfer, and R. Siegwart. Real-time monocular visual odometry for on- road vehicles with 1-point ransac. In *ICRA*, pages 4293–4299, 2009. **3**
- [26] M. Stikic, D. Larlus, S. Ebert, and B. Schiele. Weakly supervised recognition of daily life activities with wearable sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12), 12/2011 2011. **1**
- [27] S. Sundaram and W. Mayol-Cuevas. High level activity recognition using low resolution wearable vision. In *1st Workshop on Egocentric Vision, with CVPR*, June 2009. **1**
- [28] J.-P. Tardif, Y. Pavlidis, and K. Daniilidis. Monocular visual odometry in urban environments using an omnidirectional camera. In *IROS*, 2531–2538, 2008. **3**
- [29] N. Tomatis, I. Nourbakhsh, and R. Siegwart. Hybrid simultaneous localization and map building: a natural integration of topological and metric. *Robotics and Autonomous systems*, 44(1):3–14, 2003. **3**
- [30] E. A. Topp and H. I. Christensen. Detecting region transitions for human-augmented mapping. *IEEE Transactions on Robotics*, 26(4):715–720, 2010. **2**