2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Appearance-based segmentation of indoors/outdoors sequences of spherical views

Alexandre Chapoulie[1], Patrick Rives[1] and David Filliat[2]

***Abstract*— Navigating in large scale, complex and dynamic environments requires reliable representations able to capture metric, topological and semantic aspects of the scene for supporting path planing and real time motion control. In a previous work [11], we addressed metric and topological representations thanks to a multi-cameras system which allows building of dense visual maps of large scale 3D environments. The map is a set of locally accurate spherical panoramas related by 6dof poses graph. The work presented here is a further step toward a semantic representation. We aim at detecting the changes in the structural properties of the scene during navigation. Structural properties are estimated online using a global descriptor relying on spherical harmonics which are particularly well-fitted to capture properties in spherical views. A change-point detection algorithm based on a statistical Neyman-Pearson test allows us to find optimal transitions between topological places. Results are presented and discussed both for indoors and outdoors experiments.**

I. Introduction

Navigating in large scale, complex and dynamic environments is a challenging task for autonomous mobile robots. Reliable representations able to capture metric, topological and semantic aspects of the scene have to be built for supporting path planing and real time motion control algorithms [14]. It is usual to define three levels of representation as illustrated in fig. 1. Metric representation is used at the control level in the design of trajectory tracking algorithms [4]. Topological representation captures the environment accessibility properties in a graph structure and provides a first level of abstraction allowing complex navigation tasks in large scale environments [21]. Semantic representation consists in adding information about the places represented by nodes in the graph used at the topological level. The semantic information can be basically the name of a place [16] or its main characteristic such as office or corridor [24]. The added information can also refer to objects presence or other kind of information linked to the place. This level, with a higher degree of abstraction, allows us to specify context-based navigation tasks in terms of queries [7].

In [11], we addressed metric and topological representation levels thanks to a multi-cameras system onboard a man-driven car which allows building of dense visual maps of large scale 3D environments. As in Google Street View [23], the map is composed of a set of locally accurate spherical panoramas (fig. 2) built online along the car trajectory. The spherical views are related by 6dof poses graph estimated using a direct multi-views registration technique [12].

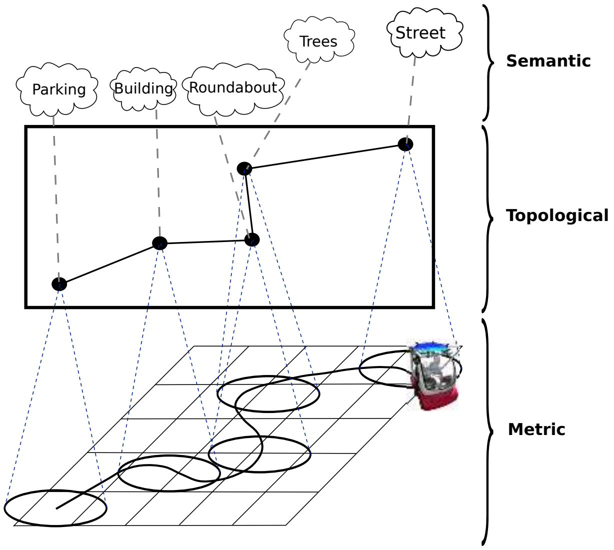

Fig. 1. Navigation-based representation

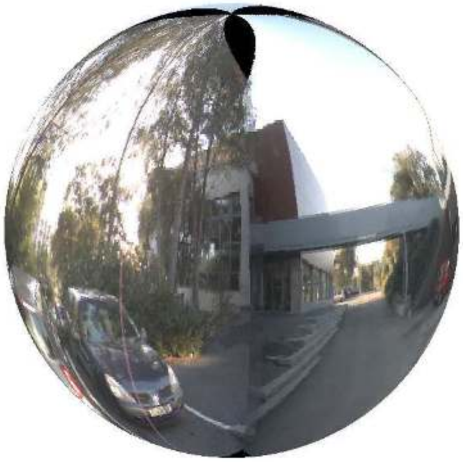

Fig. 2. Example of spherical view (Inria Campus Dataset).

The work presented here is a further step toward a semantic representation of the scene. We aim at detecting changes in the scene structural properties (such as textures, appearance, frequency and orientation of the straight lines, curvatures, repeated patterns) during navigation. A *place*, in this work, is therefore associated to a segment of the robot trajectory where the scene is sufficiently self similar, *i.e.* has

[1] INRIA Sophia Antipolis - Méditerranée, 2004 route des Lucioles - BP 93, 06902 Sophia Antipolis, France firstname.lastname@inria.fr

[2] ENSTA ParisTech, 32 Boulevard Victor, 75739 Paris, France david.filliat@ensta-paristech.fr

978-1-4673-6357-0/13/$31.00 ©2013 European Union

the same structural properties extracted from the spherical views. The main advantage of this definition is that it fits both to indoor and outdoor environments in order to partition the topological graph in terms of meaningful places. Such partition also provides advantages such as increasing loop closure algorithms efficiency [10] and can be viewed as a first step to environment semantic labeling.

In [3], we presented preliminary results where the structural properties were estimated using a global descriptor called GIST specially modified to deal with spherical images. Given our place definition, GIST appears more adapted than local descriptors like SIFT used in [17] and [25]. Without additional constraints, local descriptors have difficulty to represent the environment global consistency. Since it has been introduced [15], GIST has been used multiple time in image-based learning algorithms and in robotics for place recognition and loop closure detection [13] or for indoor region classification [18]. Despite these good properties, GIST is not well adapted to encompass the spherical representation richness because sphere spatial periodicity is partially lost. In this paper, we propose a novel representation relying on spherical harmonics which are particularly well-fitted to capture the structural properties in spherical views.

In the following, section 2 presents the representation based on spherical harmonics. Section 3 is devoted to the detection of statistical changes in the scene structural properties. Experimental results for indoor and outdoor environments are provided in section 4. The proposed method is discussed in section 5.

II. Spherical harmonics

Spherical harmonics are similar to the 2D Fourier transform but defined on the sphere surface and take complete advantage of the spherical representation. Noticeably, the complete spatial periodicity of the sphere is integrated into the spherical harmonics computation. They have already shown their usefulness in the domain of robotics for localization [5] and for visual odometry [9]. Spherical harmonics will be used here to define a new scene structure descriptor.

A. Definition

In this paper, we only detail the application of spherical harmonics to our problem. Further mathematical details about spherical harmonics can be found in [2], [1], [8].

The unit sphere S^2 included in $\mathbb{R}^3$ is parametrized using spherical coordinates. An element η of S^2 is written:

$$\eta = \left[cos(\theta)sin(\phi), sin(\theta)sin(\phi), cos(\phi)\right]^T \qquad (1)$$

The spherical harmonics are defined by:

$$Y_l^m(\eta) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^{|m|}\left(cos\left(\phi\right)\right) e^{jm\theta} \qquad (2)$$

with $l \in \mathbb{N}$ and $|m| \leq l$ where l is the band number corresponding to a frequency and m is an orientation parameter. P_l^m corresponds to the associated Legendre polynomials with $x \in [-1, 1]$ such that:

$$P_l^m(x) = \frac{(-1)^m(1-x^2)^{m/2}}{2^l l!}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l \qquad (3)$$

Every function defined on the sphere surface can be decomposed in a sum of spherical harmonics as follows:

$$f = \sum_{l\in\mathbb{N}}\sum_{|m|\leq l} f_l^m Y_l^m \qquad (4)$$

The f_l^m coefficients are obtained from a function f by:

$$f_l^m = \int_{\eta\in S^2} f(\eta)\overline{Y_l^m(\eta)}d\eta \qquad (5)$$

If $f_l^m = 0$ for all $l > L$, f is said to be band limited with a bandwidth L. The coefficients set f_l^m is called the spherical Fourier transform or the spectrum of f. The first five spherical harmonics bands are displayed in fig. 3.

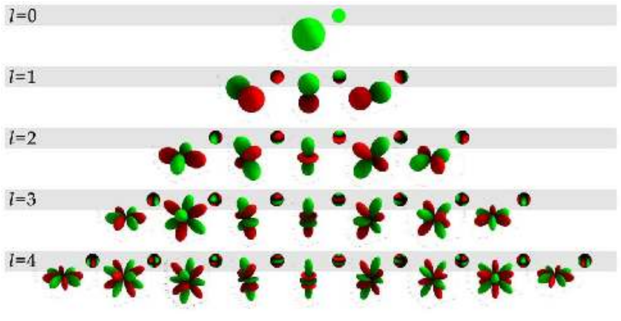

Fig. 3. The first five spherical harmonics bands are presented as unsigned spherical functions from the origin and by color on the unit sphere. Green corresponds to positive values and red to negative values. (From [8])

Due to the integral, f_l^m coefficients exact computation can be very time consuming. While it exists the fast Fourier transform, there exists a fast method to compute those coefficients, based on the Monte Carlo integration, precomputed tables and the properties of the associated Legendre polynomials. This method is widely used in computer graphics for real-time lighting rendering. Further details can be found in [8].

B. Spherical harmonics as environment structure description

Assuming that environment structure information is contained in the spherical image frequencies, pixel intensities can be chosen as the samples $f(x_i)$ values of the function f. Spherical harmonics being a frequency description of the spherical image, we propose to directly use the spectrum as a structure descriptor. Frequency information corresponds to band number l and orientation information to parameter m (the higher l is, the higher the frequency is, see fig. 3). The spectrum coefficients f_l^m are stacked into a vector which constitutes the global structure descriptor.

The number of bands used is an important parameter. In the case of the 2D discrete Fourier transform, the spectrum size is constrained by the image size. In the case of the spherical harmonics, nothing constraints the required number of bands. The number of coefficients follows a square function of the number of bands. The descriptor size is $S_d = l^2$. In fig.3, $l = 5$ and we have $l^2 = 25$ coefficients.

In computer graphics, only three bands are used due to an exponential attenuation in bands of higher frequencies [8]. For our study, there is no such attenuation and it is hard to determine the required number of bands. In [5], precise localization is achieved using only the first five bands. While we seek a global description of the environment, the first five bands should guarantee a sufficient information.

III. Change-point detection algorithm

A. Hypotheses and assumptions

According to our place definition as a set of positions from which environment structure is similar, we aim to detect the significant changes in the global descriptor value along the sequence of spherical views. This can be viewed as novelty detection as used in [19] or [20] for vehicle safeguarding or as change-point detection as used in [17] and [16] for landmark detection and place labelling. Change-point detection is based on hypothesis testing:

- Null hypothesis H_0 is the normal situation in which the observed parameters stick to the previous model.
- Alternate hypothesis H_1 is the alternate situation where parameters vary from the previous model.

Change-point detection algorithm evaluates the monitored parameters and determines when a switch occurs from hypothesis H_0 to hypothesis H_1.

Let us assume a set of independent input observations:

$$X_1, X_2, ..., X_{\tau-1}, X_\tau, ..., X_t \tag{6}$$

Assume that the input observations $X_1, ..., X_{\tau-1}$ are independent random variables with a probability density function $f_0(X_j)$, while the observations $X_\tau, ...$ are independent random variables with a probability density function $f_1(X_j)$. Let us assume that f_0 is the probability density function under hypothesis H_0 and f_1 under H_1. Suppose we have $X_1, ..., X_t$ observations up to an instance t and we test the above hypotheses for these observations. The likelihood ratio (eq. 7) indicates whether the value X_j mostly belongs to f_1 or f_0.

$$s_j = ln \frac{f_1(X_j)}{f_0(X_j)} \tag{7}$$

The Neyman-Pearson lemma conducting a simple hypothesis test, as used in [22], defines the uniformly most powerful test as the one rejecting the null hypothesis H_0 whenever:

$$S_\tau^t = \sum_{j=\tau}^{t} ln \frac{f_1(X_j)}{f_0(X_j)} = \sum_{j=\tau}^{t} s_j > \nu \tag{8}$$

The above equation yields to the simple hypothesis test:

$$t_c = min\{t : \underset{0 \le \tau \le t}{\arg\max} \sum_{j=\tau}^{t} ln \frac{f_1(X_j)}{f_0(X_j)} > \nu\} \tag{9}$$

where ν is the threshold controlling the detection sensitivity. $\underset{0 \le \tau \le t}{\arg\max} \sum_{j=\tau}^{t} ln \frac{f_1(X_j)}{f_0(X_j)} > \nu$ returns the instant τ giving the maximum of dissimilarity between f_0 and f_1. t being the current instant, t_c will be either t leading to no change-point detection or τ which is the exact change-point instant.

This algorithm gives the exact change-point instants whereas it needs a delay to evaluate the probability density function f_1. The computation time is very low for a small t but increases rapidly with the number of observations. No assertions are done concerning H_0 and the probability density functions f_0 and f_1 always need to be estimated for all the change-points τ tested over all observations.

Let's assume the density functions under each hypothesis, *i.e.* f_0 and f_1, follow a multivariate normal distribution:

$$f_0 \sim \mathcal{N}(\mu_0, \Sigma_0) \qquad f_1 \sim \mathcal{N}(\mu_1, \Sigma_1) \tag{10}$$

As each hypothesis is characteristic of one topological place, density functions characterize the structural parameters of topological places. The mean vector represents the most probable structural parameters set. The covariance matrix represents the parameters distribution tolerance inside a topological place. Two matters arise concerning the distributions parameters estimation:

- Sufficient number of samples are necessary to insure well conditioned density function estimation and in particular the covariance matrix semi-definite positiveness property.
- Density function estimation requires identically and independently distributed samples (*i.i.d*). Independence is assumed due to independent input observations assumption from Neyman-Pearson lemma. Approximate constant distance interval gathering (constant time gathering with minimal distance between samples condition) allows approximate identical distribution. This simple method avoids accumulation at low or null speed.

B. Online application

As explained previously, the algorithm rapidly becomes time consuming and only one change-point detection is possible for a complete set of input observations. In order to alleviate those limitations, we introduce a fixed size sliding window over the signal made up of the input observations (fig. 4). First half of the sliding window corresponds to normal hypothesis H_0 while second half corresponds to alternate hypothesis H_1. Change-point hypothesis is then tested only at the sliding window center. Each time the robot acquires a new observation, the signal is expanded with a new input. The sliding window always encompasses the N last input observations. Older observations, already analysed, are forgotten. We finally obtain an approximation (due to non complete signal observation) of the exact change-point.

This simple trick brings many advantages. The most obvious ones are constant time change-point detection and dynamic signal analysis leading to an inline algorithm. Moreover, one of the most important is multiple hypothesis testing. This last one allows to have many change-points over the signal contrarily to the original Neyman-Pearson algorithm formulation.

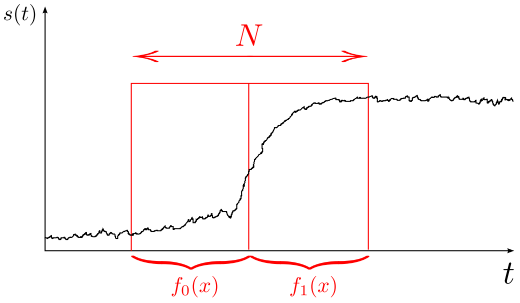

Fig. 4. Sliding window used in the estimation process.

Considering hypotheses about the density functions and the sliding window trick, the Neyman-Pearson final equation results in:

$$S_{\tau}^{t} = \frac{N}{4} ln \left(\frac{|\Sigma_0|}{|\Sigma_1|} \right) + \frac{N}{4} \left(\mu_0^T \Sigma_0^{-1} \mu_0 + \mu_1^T \Sigma_1^{-1} \mu_1 - 2 \mu_0^T \Sigma_0^{-1} \mu_1 \right) + \frac{1}{2} \sum_{j=t-N/2}^{t} \left(X_j^T \left(\Sigma_0^{-1} - \Sigma_1^{-1} \right) X_j \right) \qquad (11)$$

The equation contains three terms:

- First term is linked to distribution spreads. The term is canceled for equal spreads.
- Second term approximately corresponds (because of impossible factorization) to the squared difference between distribution means.
- Last term is the sum of the squared observations weighted by the spread difference between the density functions. The term is canceled for equal spreads.

As stated before, we can observe that the equation computes a value linked to the difference between two distributions. The greater the difference is, the higher the value is. In our case, this leads to change-point detection indicating a change in the structural parameters, which corresponds to a transition between two topological places.

An example of signal obtained with equation 11, made up of the change-point values, is displayed in fig. 5. The signal is filtered in the time domain with a simple Gaussian filter (parameters: $\mu = 0$, $\sigma = N/10$) in order to reduce the signal noise. Peak detection mechanism relies on peak magnitude relatively to the minima flanking the peak. Threshold ($\nu = 0.4$) is then used on the peak amplitude and not on the peak maximum value. This results in a peak detection less sensitive to noise.

Considering the density function estimation constraints aforementioned, the sliding window has to be sufficiently large for a correct estimation. For the experiments, the size is of 80 observations. As the minimal distance between two samples is 0.015m, the sliding window spatial size is 1.2m. Each density function is then estimated over a distance of 0.6m. These values satisfy the requisites for density estimation but has consequences on the experiment as two change-points cannot be closer than 0.6m for detection. This distance is a reasonable trade-off between minimal environment size for structural parameters extraction and minimal detectable topological place. For environments changing slowly, the window can be larger.

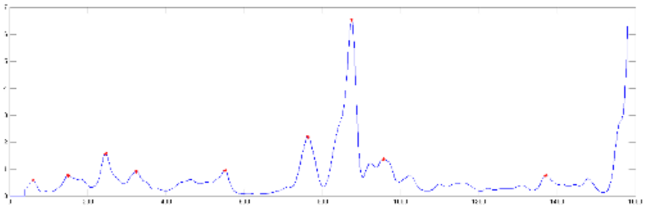

Fig. 5. Sample signal obtained with the change-point detection algorithm combined with spherical harmonics approach for structural parameters description. Detected peaks are marked with red dots.

IV. Experimental Results

This section presents experimental results for topological segmentation in indoor and outdoor environments. Testing different kind of environment aims to show the method is generic and robust to context change. Using various kind of camera for spherical view acquisition furthermore highlights the generic spherical concept. The indoor experiment was realized in the Robotic Hall at INRIA Sophia Antipolis using a Neobotix MP-500 platform equipped with a paracatadioptric camera. In the outdoor experiment, a man-driven vehicle equipped with the multi-cameras system described in [11] was used. The trajectory was about 600 meters across the INRIA Sophia Antipolis research center.

The whole code is written in Matlab without being specifically optimized. Spherical harmonics spectrum computation requires 290ms using the implementation described above (the sphere is sampled with 62500 samples uniformly distributed). The change-point detection algorithm runs in 10ms. The complete algorithm then runs inline in about 300ms (acquisition up to 3.3Hz). However, the spherical harmonics spectrum code is highly parallelizable and might take great advantage of a C/C++ parallel implementation.

A. Indoor experiment analysis

Figure 6 presents the robot trajectory and the detected change-points. It is first interesting to notice that all change-points correspond to important structure variations such as doorsteps or room volume variation (*i.e.* passing from a nook to a more open space). The trajectory in the wide space is very little segmented.

The easiest way to validate a topological place segmentation algorithm is to consider the doorsteps case. This case is illustrated by images 2680, 3480, 5328, 10455, 11954 and 12322 where change-points are precisely localized at doorsteps. The examples illustrated by images 996, 1401 and 2044 correspond to room volume variations. Image 996 and 1401 show when the robot comes from a narrow space to a wider space. Image 2044 shows the opposite case when the robot leaves a wide environment to enter a quite narrow place similar to a corridor. Images 6376 and 6624 correspond to the detection of changes in the objects present in the

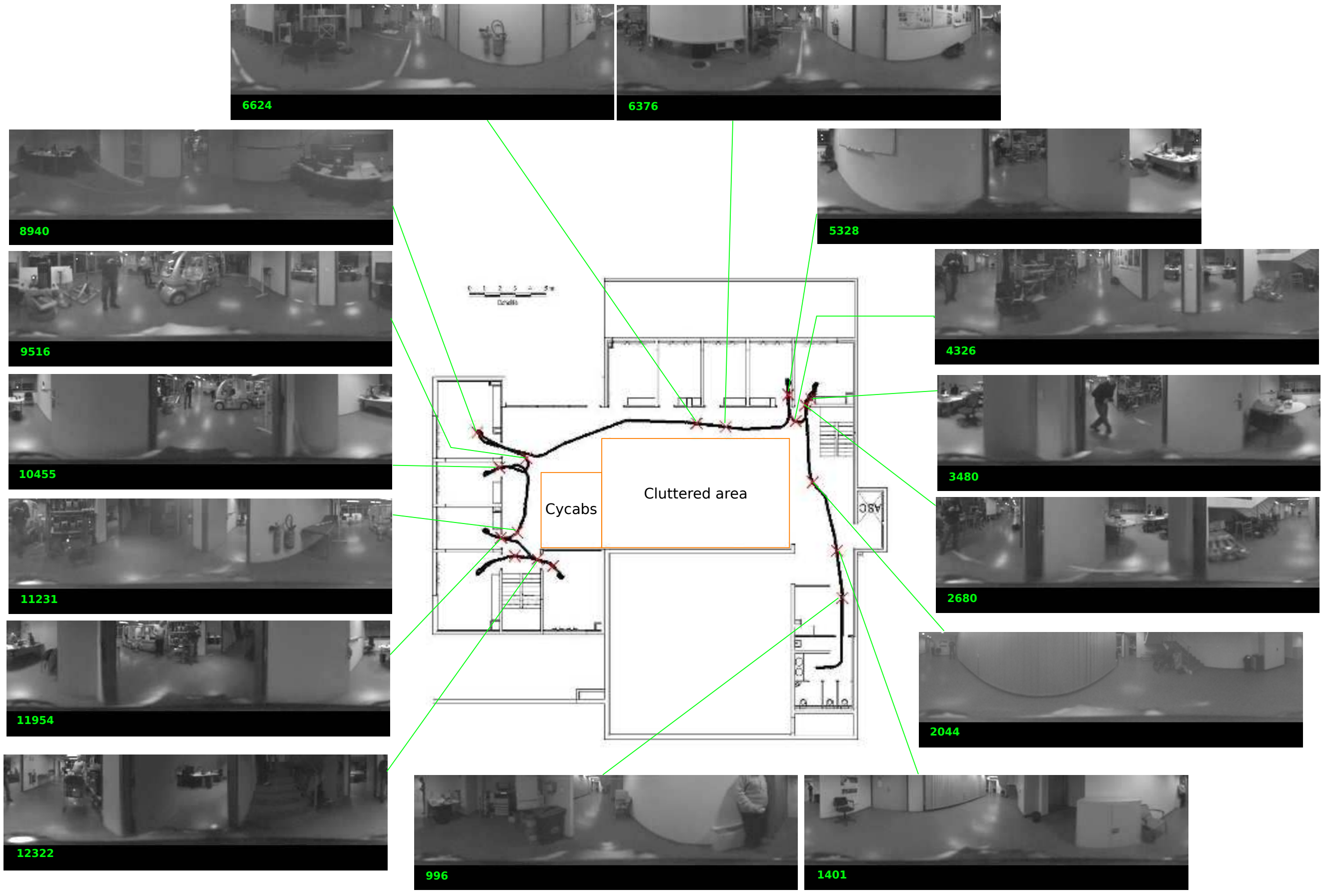

Fig. 6. Indoor trajectory inside the Robotic Hall. Detected change-points are marked with red crosses.

environment. The images (9516, 11231) define the space between the wall and the electric vehicles.

All those aforementioned change-points are relevant and are very significant considering the topological place definition we gave. There are however some false and missing change-points. Concerning the false change-points, one is illustrated by image 8940 in the upper left office. This change-point is detected while the robot was turning around, we suppose the problem is due to strong illumination variations in the images caused by the automatic shutter of the camera. Conversely, a change-point which should be detected is missing at the entrance of the same office.

B. Outdoor experiment analysis

The results are shown in fig. 7. The parking areas are clearly identified in (630, 842) and in (1035,1450). In the last case, it is interesting to notice that this parking is long enough and has a significant curve to prevent mutually seen features from the beginning and the end of the parking. This demonstrates that detection is linked to the intrinsic structure associated to the parking area and not to the observation of same objects along the sequence of views. This behavior perfectly fits what we aimed by giving an original topological place definition. Globally, the changes between the buildings and the vegetation areas are also well detected (229, 325, 403). A change-point, image 842, occurs when the vehicle crosses under a sidewalk and discovers a new area.

V. Conclusion and Future Work

We have presented an new method to cluster images into significant topological places. A place is defined as a segment of trajectory where the structural properties extracted from spherical views are sufficiently self similar. Place characterization is made by a global descriptor given by the spherical harmonics spectrum. The segmentation algorithm relies on an efficient change-point detection based on multi-hypothesis testing and allowing constant time computation. Results are very satisfying for both indoor and outdoor environments.

While the results are very good, the algorithm still shows some limitations. As descriptors are based on appearance frequencies, when the robot approaches walls, frequencies become lower and a new topological place is defined.

For future work, we plan to improve our algorithm robustness to illumination condition following [6] and its rotation independence. The algorithm presents a certain robustness to rotation due to the sliding window reducing the environment sensed, but the spherical harmonics spectrum is not independent to any rotation. De-rotation mechanism can be applied as rotations can be estimated from spectra.

In a longer term, the segmentation algorithm could be coupled with a loop closure detection algorithm in order

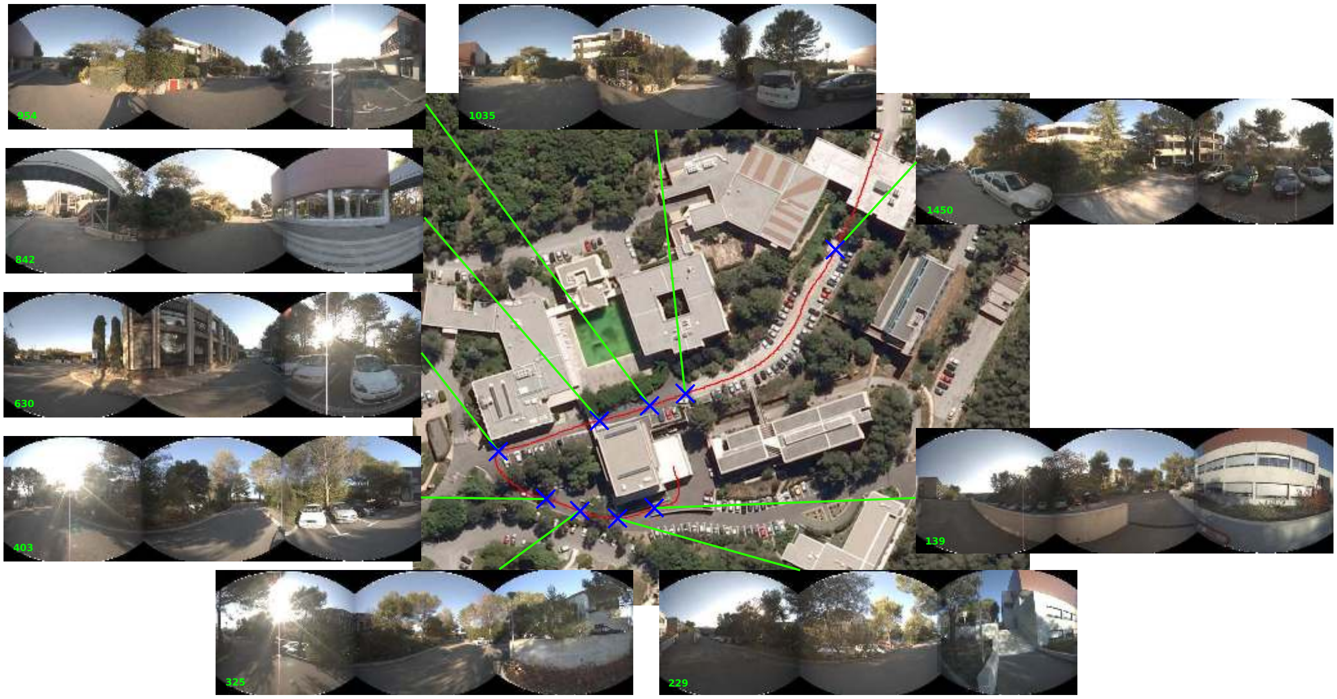

Fig. 7. Outdoor experiment in the INRIA campus with the spherical harmonics feature. Detected change-points are marked with blue crosses.

to improve change-point localization stability and with a semantic level by adding place classification and labelling. Finally, experiments with drones could test rotation independence and validate the generic approach elaborated.

REFERENCES

[1] T. Bülow. Multiscale image processing on the sphere. In Luc J. Van Gool, editor, *DAGM-Symposium*, volume 2449 of *Lecture Notes in Computer Science*, pages 609–617. Springer, 2002.

[2] T. Bülow and K. Daniilidis. Surface representations using spherical harmonics and gabor wavelets on the sphere, 2001.

[3] A. Chapoulie, P. Rives, and D. Filliat. Topological segmentation of indoors/outdoors sequences of spherical views. In *IEEE/RSJ International Conf. on Intelligent Robots and Systems (IROS)*, 2012.

[4] J. Courbon, Y. Mezouar, and P. Martinet. Autonomous navigation of vehicles from a visual memory using a generic camera model. *Trans. Intell. Transport. Sys.*, 10(3):392–402, September 2009.

[5] H. Friedrich, D. Dederscheck, K. Krajsek, and R. Mester. View-based robot localization using spherical harmonics: Concept and first experimental results. In *DAGM-Symposium*, volume 4713 of *Lecture Notes in Computer Science*, pages 21–31. Springer, 2007.

[6] H. Friedrich, D. Dederscheck, M. Mutz, and R. Mester. View-based robot localization using illumination-invariant spherical harmonics descriptors. In *VISAPP (2)*, pages 543–550, 2008.

[7] C. Galindo, J.-A. Fernandez-Madrigal, J. Gonzlez, and A. Saffiotti. Robot task planning using semantic maps. *Robotics and Autonomous Systems (RAS*, 56(11):955–966, 2008.

[8] R. Green. Spherical Harmonic Lighting: The Gritty Details. *Archives of the Game Developers Conference*, march 2003.

[9] H. Hadj-Abdelkader, E. Malis, and P. Rives. Spherical image processing for accurate visual odometry with omnidirectional cameras. In *8th Workshop on Omnidirectional Vision, (Omnivis)*, Marseille, France, October 2008.

[10] H. Korrapati, J. Courbon, and Y. Mezouar. Topological mapping with image sequence partitioning. In *Intelligent Autonomous Systems 12*, volume 193 of *Advances in Intelligent Systems and Computing*, pages 143–151. Springer Berlin Heidelberg, 2013.

[11] M. Meilland, A.I. Comport, and P. Rives. A Spherical Robot-Centered Representation for Urban Navigation. In *IEEE/RSJ International Conf. on Intelligent Robots and Systems (IROS)*, 2010.

[12] M. Meilland, A.I. Comport, and P. Rives. Dense visual mapping of large scale environments for real-time localisation. In *IEEE/RSJ International Conf. on Intelligent Robots and Systems (IROS)*, 2011.

[13] A. C. Murillo, G. Singh, J. Košecká, and J. J. Guerrero. Localization in urban environments using a panoramic gist descriptor. *Robotics, IEEE Transactions on*, 29(1):146–160, 2013.

[14] P. Newman, G. Sibley, M. Smith, M. Cummins, A. Harrison, C. Mei, I. Posner, R. Shade, D. Schrter, L. Murphy, W. Churchill, D. Cole, and I. Reid. Navigating, recognising and describing urban spaces with vision and laser. *The International Journal of Robotics Research*, 2009.

[15] 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, May 2001.

[16] A. Ranganathan. PLISS: Detecting and Labeling Places Using Online Change-Point Detection. In *Robotics: Science and Systems VI*, 2010.

[17] A. Ranganathan and F. Dellaert. Bayesian surprise and landmark detection. In *ICRA*, pages 2017–2023. IEEE, 2009.

[18] A. Rituerto, A.C. Murillo, and J.J. Guerrero. Semantic labeling for indoor topological mapping using a wearable catadioptric system. *Robotics and Autonomous Systems*, (0):–, 2012.

[19] A. J. Smola, L. Song, and C. H. Teo. Relative novelty detection. *Journal of Machine Learning Research - Proceedings Track*, 5:536–543, 2009.

[20] B. Sofman, J. A. Bagnell, and A. Stentz. Anytime online novelty detection for vehicle safeguarding. In *ICRA*, pages 1247–1254. IEEE, 2010.

[21] N. Tomatis, I. R. Nourbakhsh, and R. Siegwart. Hybrid simultaneous localization and map building: a natural integration of topological and metric. *Robotics and Autonomous Systems (RAS*, 44(1):3–14, 2003.

[22] G. Tsechpenakis, D. N. Metaxas, C. Neidle, and O. Hadjiliadis. Robust online change-point detection in video sequences. In *Proceedings of the Conf. on Computer Vision and Pattern Recognition Workshop*, pages 155–, 2006.

[23] L. Vincent. Taking online maps down to street level. *Computer*, 40:118–120, 2007.

[24] J. Wu, H. I. Christensen, and J. M. Rehg. Visual place categorization: Problem, dataset, and algorithm. In *IROS*, pages 4763–4770. IEEE, 2009.

[25] Z. Zivkovic, B. Bakker, and B. Krose. Hierarchical map building using visual landmarks and geometric constraints. In *Intelligent Robots and Systems (IROS). IEEE/RSJ International Conference on*, pages 2480 – 2485, August 2005.