

Reliable Topological Place Detection in Bubble Space

Hakan Karaoguz and H. Işıl Bozma

Abstract—This paper introduces a novel approach to topological place detection. The approach is based on previously proposed bubble space representation - where all sensory features and their relative S^2 - geometry are encoded in a manner that is implicitly dependent on robot pose. Its novelty is that ensuring sensory data reliability is integrated with place detection. This is achieved via checking for informativeness, coherence and plenitude using only the bubble space representation of the incoming sensory data. The stringency of these checks is controllable via a set of associated parameters. Experimental results with benchmark datasets indicate correct detection rates comparable to state-of-the-art approaches in place detection. Furthermore, the detected places can then be immediately used to generate the nodes in topological maps.

I. INTRODUCTION

Topological maps represent the continuous world as a finite set of nodes ("places") connected by edges ("paths") [1]. The nodes of the graph generally represent distinctive locations while edges represent transitions between these nodes [2]. Its definition varies between being as a human recognizable structure [3] or a junction or transition region [4]. Place detection aims at the generation of these nodes and thus is critical to topological maps. As the robot navigates through the environment, place detection requires segmenting the sensory streams into pieces corresponding to distinct places [4]. The goal is to have reliable place detection [4], [5].

This problem has been addressed via different perspectives. In all, place detection occurs when there is a significant change in sensory readings. Approaches can be categorized depending on whether odometric and/or metric data is used or not. In the former, place detection is based on geometric reasoning. For example, places are defined as intersections on Voronoi graphs obtained from metric maps [4]. Bayesian reasoning with laser measurements is used to detect places, crossings and doorways [6]. In another work, geometric features are used in an indoor environment for door-passage-independent detection of transitions between places [7]. However, the reliability of the metric data may not be always ensured [1].

Alternatively, appearance based approaches aim to detect places only by using visual sensory information without any metric information. One common approach has been to formulate the problem as a scene change detection or image sequence partitioning problem and resort to various video processing techniques [3], [8]. There are three main approaches for appearance based place detection. These are image-based, feature-based and histogram-based approaches.

In image-based approaches, the pixel value differences or statistical properties between consecutive images are processed in order to detect scene changes [9]. In feature-based approaches, various features such as edges or SURF features are used to find the similarity between perspective image frames [10] or omnidirectional images [8]. In histogram-based approaches, images are compared after obtaining histogram representations [11]. Furthermore, these approaches have been extended to machine vision techniques such as, optical flow information is used to detect and describe the changes in the environment [12].

The problem with most of these approaches is that they have difficulties in distinguishing between places that are known, but viewed with varying conditions (i.e. illumination, viewpoint and/or occlusions) and unknown places. There has been numerous work that have addressed this issue. One approach is to detect challenging instances in the recognition module where the detection is based on confidence measures associated with label assignment process [13]. Another approach is to use statistical approaches to changepoint detection [14], [15]. An online Bayesian approach that detects changes to model parameters is used to segment the image stream [16]. Interestingly, most work assume that incoming sensory data can be used immediately. However, if the data is problematic, this may conceivably decrease the reliability of place detection.

This paper presents a novel approach to place detection in topological maps. The robot is assumed to acquire its data by a perspective camera or Kinect camera in temporal continuity¹ [17]. The robot has neither any environmental knowledge nor uses odometric data. Our approach is based on bubble space - where all sensory features and their relative S^2 - geometry are encoded in a manner that is implicitly dependent on robot pose via bubble surfaces and bubble descriptors [18]. The novelty of this approach is that as part of place detection, sensory data is checked for reliability. This is achieved by checking for informativeness, coherence and plenitude using only bubble space representation of the incoming sensory data. The stringency of these checks is controllable via a set of associated parameters. Experimental results demonstrate that as such the robot is able to detect places reliably with comparative performance with the state-of-the-art approaches. Furthermore, the detected places can then be immediately used for node generation in topological maps and for place recognition. The outline of this paper is as follows: First, bubble space is explained briefly in Section II

Intelligent Systems Laboratory, Electrical and Electronic Engineering Bogazici University, Bebek 34342 Istanbul Turkey
 hakan.karaoguz@boun.edu.tr

¹As such, its field of view is very limited - for example as compared to omni-directional images.

for completeness. Next, data reliability is explained in Section III. Section IV presents the place detection algorithm. Experimental results with the two benchmark ImageCLEF 2013 RGB-D dataset and VPC 2009 dataset are presented in Section V. The paper concludes with a brief summary and comments regarding future directions.

II. BUBBLE SPACE

In this section, bubble space representation is reviewed briefly for completeness. The interested reader is referred to [18] for details. Consider a robot at base $x \in \mathcal{X}$ at time t defined as $x = [c \ \alpha]^T$ with position $c \in R^2$ and heading $\alpha \in S^1$. The base space is $\mathcal{X} \subset R^2 \times S^1$. The robot sensor state space $\mathcal{F} \subset S^2$ refers all viewing directions. The bubble space $\mathcal{B} = \mathcal{X} \times \mathcal{F}$ is an egocentric representation with $b \in \mathcal{B}$ defined as $b = [x \ f]^T$. For each viewing direction $f \in \mathcal{F}$, it derives a set of features \mathcal{L} with $|\mathcal{L}| = N_l$. Let $q(b, t) = [q_1(b, t), \dots, q_{N_l}(b, t)]^T$ denote the observation vector that consists of all individual responses. Now, for each feature $i \in \mathcal{L}$, visualize the robot to be surrounded by an hypothetical spherical surface - referred to as bubble surface $B_i(x, t)$ - that is deformed at each f by an amount that is dependent on the feature response in that direction. Mathematically, each bubble surface $B_i(x, t)$ is a deformed sphere embedded in R^3 with an intrinsic parametrization:

$$B_i(x, t) = \left\{ \left[\begin{array}{c} f \\ \rho_i(b, t) \end{array} \right] \mid \forall f \in \mathcal{F} \text{ and } b = [x^T \ f^T]^T \right\} \quad (1)$$

where $\rho_i : \mathcal{B} \times R^{\geq 0} \rightarrow [\rho_{min}, \rho_{max}]$ is a Riemannian metric that encodes the response to the i^{th} feature. The parameters $0 < \rho_{min} < \rho_{max}$ define the lower and upper bounds on individual feature values. The initial bubble surface $B_i(x, 0)$ is defined as a S^2 sphere with radius $\rho_0 \in [\rho_{min}, \rho_{max}]$ - namely $\rho_i(b, 0) = \rho_0$. The surface $\rho_i : \mathcal{B} \times R^{\geq 0} \rightarrow [\rho_{min}, \rho_{max}]$ can be explicitly represented by the double Fourier series as [19]:

$$\rho_i(b, t) = \sum_{m=0}^{H_1} \sum_{n=0}^{H_2} \lambda_{mn} z_{xi,mn}^T(t) e_{mn}(f)$$

with predefined parameters λ_{mn} . For each (m, n) , the vector $e_{mn}(f) \in R^4$ consists of an orthonormal set of trigonometric basis functions. The corresponding vector $z_{xi,mn}(t) \in R^4$ of coefficients of double Fourier series associated with base point x at time t is defined as:

$$z_{xi,mn}(t) = [\eta_{xi,mn}(t) \ \beta_{xi,mn}(t) \ \mu_{xi,mn}(t) \ \nu_{xi,mn}(t)]^T \quad (2)$$

where letting $f \in \mathcal{F}$ as $f = [f_1 \ f_2]^T$,

$$z_{xi,mn}(t) = \frac{1}{\pi^2} \begin{bmatrix} \int_0^{2\pi} \int_0^{2\pi} \rho_i(b, t) \cos(mf_1) \cos(nf_2) df_1 df_2 \\ \int_0^{2\pi} \int_0^{2\pi} \rho_i(b, t) \sin(mf_1) \cos(nf_2) df_1 df_2 \\ \int_0^{2\pi} \int_0^{2\pi} \rho_i(b, t) \cos(mf_1) \sin(nf_2) df_1 df_2 \\ \int_0^{2\pi} \int_0^{2\pi} \rho_i(b, t) \sin(mf_1) \sin(nf_2) df_1 df_2 \end{bmatrix}$$

Bubble descriptors are N_I -dimensional vectors where $N_I = N_l(H_1 + 1)(H_2 + 1)$ that enable holistic representation of bubble surfaces while being rotationally invariant - defined as:

$$I(x, t) = [I_{1,00}(x, t), \dots, I_{N_l(H_1)(H_2)}(x, t)]^T \quad (3)$$

where

$$I_{i,mn}(x, t) = z_{xi,mn}^T(t) z_{xi,mn}(t) \quad (4)$$

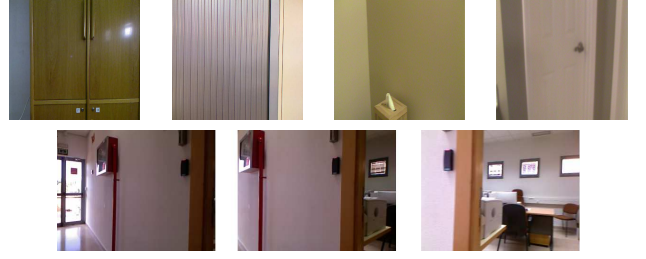


Fig. 1. Examples of unreliable sensory data. Top: Uninformative data ImageCLEF 2013 (first two images) and VPC 2009 (last two images); Bottom: Incoherent data from 3 consecutive base points.

III. SENSORY DATA RELIABILITY

There are several reasons as to why the incoming sensory data may not be reliable. In our work, we identify three primary factors²:

- **Informativeness:** The sensory data collected at a particular base point may not be informative due to problematic environmental conditions (illumination, viewpoint). Examples of low illumination conditions or viewpoints close to a wall are as shown in Fig. 1(top).
- **Coherency:** Sensory data from two or more consecutive base points may not be coherent. A sample case of incoherent sensory data from 3 consecutive base points is as shown in Fig. 1(bottom). The incoherency may manifest itself as abrupt changes of short or extended duration. The former may arise when there are random and temporary acquisition problems while the latter will arise in transition between two different places. As such, the type of incoherency also needs to be determined.
- **Plenitude:** The number of base points from which sensory data is available may be too small. Recent findings suggest that recognition performance using single images or short image sequences tends to be poor while improving vastly as with the size of image data used [20].

In all cases, data collected from the associated base point needs to be ignored.

A. Data Informativeness

Data informativeness requires a measure to assess whether an incoming sensory data is semantically rich or not. In particular, the variation in statistical properties of feature values can serve as an indicator of informativeness. The bubble surfaces $B_i(x, t)$ allow the computation of such statistical properties. In particular, the average surface deformation and variance can be easily computed. By definition, the Fourier

²Of course, if additional sources of unreliability are determined, this list needs to be extended accordingly.

coefficient $\eta_{xi,00}$ corresponds directly to average surface deformation.

$$\eta_{xi,00} = \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{2\pi} \rho_i(b, t_k) \cos(mf_1) \cos(nf_2) df_1 df_2 \quad (5)$$

The surface variance σ_{xi} can be computed as:

$$\sigma_{xi} = \int_0^{2\pi} \int_0^{2\pi} (\rho_i(b, t_k) - \eta_{xi,00})^2 df_1 df_2 \quad (6)$$

If the sensory data from a base point x is uninformative, the bubble surface statistics as measured by the average deformation or surface variance will be low. Under these conditions, the sensory data needs to be ignored. Hence, the informativeness of the associated data can be defined by the binary valued function $\varsigma : \mathcal{X} \rightarrow \{0, 1\}$

$$\varsigma(x) = \begin{cases} 0 & \eta_{xi} < \tau_{i1} \\ 0 & \sigma_{xi} < \tau_{i2} \\ 1 & \text{otherwise} \end{cases} \quad (7)$$

where the parameters τ_{i1} and τ_{i2} are a priori selected informativeness thresholds. The index $i \in \mathcal{L}$ represents the feature used. Sensory data from a particular base point x is used if and only if $\varsigma(x) = 1$. For example, the visual data can be checked for the illumination conditions. This can be accomplished easily with bubble surface associated with the intensity feature. If its average deformation is low, this indicates that robot is looking at a very dark scene which will not provide much information. Likewise, if the bubble surface variance is low, the robot is probably looking at a very evenly illuminated scene such as a door or wall which will not provide valuable information either. Similarly, the bubble surface associated with depth feature can be used in detecting whether the robot's viewpoint is blocked or not. If the average deformation of the corresponding bubble surface is low, this means the robot is very close to a wall or object that will yield poor sensory information. A low bubble surface variance is an indication that the robot is standing in front of a very flat surface such as a wall.

B. Coherency

Coherency is based on data consistency. It is measured by comparing the similarity of the data obtained from two consecutive base points x_k and x_{k+1} . In particular, our similarity measure is based on comparing the respective bubble descriptors $I(x_k)$ and $I(x_{k+1})$ using a distance metric d . In our approach, $d = \chi^2$ distance metric is used. A low similarity value is an indicator of incoherency. A coherency threshold τ_3 is used for this purpose. Any pair of consecutive data with similarity lower than τ_3 is thought to be incoherent. A temporal window is initiated when a robot detects low similarity - if one has not been started already. When a temporal window is initiated due to a detected incoherency, the sensory data associated with a number of succeeding base points needs to be checked in order to detect transitions robustly. This is achieved by keeping track the dissimilarities in the temporal window. The coherency parameter τ_n defines the number of succeeding base points that will be checked

once a temporal window is started. If there is at least one base point with incoherent data in the next τ_n base points, then the temporal window is extended to include that base point associated with incoherent data as $[t_k - \delta t_k, t_k]$ where $t_k - \delta t_k$ is the start time and t_k is the last time an incoherency was detected. This process is repeated as long as an incoherency is detected. Once a temporal window terminates (there is no further incoherency in the next τ_n base points), its extent is used in order to decide whether the corresponding base points are in a transition region or not. In case the extent of base points associated with incoherent data is short, this data simply needs to be ignored. On the other hand, if the extent is long, the base points associated with the incoherent data are considered as transitions. In such cases, the robot needs to differentiate between the two places before and after the transition region. This is based on another coherency parameter τ_w . If the temporal window extent is less than τ_w , then the data associated with the corresponding base points are treated as noise and are simply ignored. In case its extent is larger than or equal to τ_w , then this region is detected as a transition region. In this case, the robot detects two places with the incoherent region in-between removed.

C. Plenitude

As the robot goes through each different place p , there will be an associated temporal window $[T_p, T'_p]$ with the following property: $\forall t_k \in [T_p, T'_p]$, the base point x_k is from the same place. A temporal window is initiated once data from consecutive base points are similar - as measured by the similarity of the corresponding bubble descriptors. The extent of this window is an indication of data size. In particular, data having temporal window extent less than a plenitude threshold τ_p is considered to be of insufficient amount and not considered.

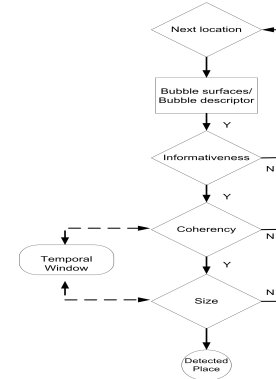


Fig. 2. Place detection overview.

IV. PLACE DETECTION OVERVIEW

Consider a navigating robot that is collecting sensory data at times t_k from base points $x(t_k)$ where $t_k < t_{k+1}$. The overview of place detection is shown in Fig. 2. First, the incoming sensory data is represented as bubble surfaces $B_i(x(t_k), t_k)$ or bubble descriptors $I(x(t_k))$. Next, the sensory data is consecutively checked for informativeness,

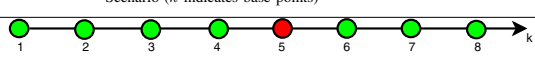


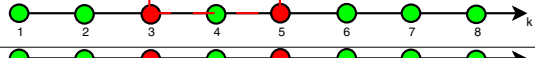
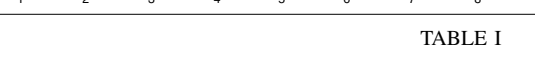
Case #	Scenario (k indicates base points)	Parameters		# Places	Detected places p_i
		τ_w	τ_n		
1		2	2	1	$p_1 = \{1 - 3, 6 - 8\}$
2		2	2	2	$p_1 = \{1 - 3\}$ $p_2 = \{7 - 8\}$
3		3	2	1	$p_1 = \{1 - 3, 7 - 8\}$
4		2	2	2	$p_1 = \{1 - 2\}$ $p_2 = \{6 - 8\}$
5		2	1	1	$p_1 = \{1 - 2, 4, 6 - 8\}$

TABLE I

COHERENCY DETECTION FOR VARYING τ_w AND τ_n VALUES. GREEN DOTS INDICATE BASE POINTS WITH SIMILAR SENSORY DATA WHILE RED DOTS INDICATE BASE POINTS HAVING DIFFERENT SENSORY DATA. EACH DETECTED PLACE p_i IS ASSOCIATED WITH BASE POINTS AS SPECIFIED.

coherency and size using these representations as explained previously. The detected places are generated accordingly. Of course, the parameter values are extremely important. The threshold values τ_{i_1} and τ_{i_2} determine the amount of data that will be directly ignored. The threshold values τ_w and τ_n affect the number of detected places. This is exemplified for sample cases in Table I for varying parameter values. In each scenario, green dots indicate base points with similar sensory data while red dots indicate dissimilar sensory data.

In the first case, only the data associated with x_5 is incoherent. Hence, the temporal window is $[t_5, t_5]$. As there is no other incoherency in the data associated with the next set of $\tau_n = 2$ base points, our temporal window size remains 0. The data associated with x_5 is simply ignored as its extent is smaller than $\tau_w = 2$. Hence, there is only one place detected with all the bubble descriptors except that associated with x_5 are taken into account. In the second case, data dissimilarity occurs at consecutive base points x_4 , x_5 and x_6 . Thus, the temporal window will start as $[t_4, t_4]$. As there is dissimilarity in the following $\tau_n = 2$ frames (at x_5 and x_6 respectively), the temporal window extends to $[t_4, t_6]$. As there is no dissimilarity afterwards, the temporal window is not extended further. Its size is equal to $\tau_w = 2$ so it is marked as a transition region with red rectangle. As the associated base points are removed from consideration, the remaining base points are partitioned into two groups. The sensory data from the first three base points are associated with one place p_1 while data associated with base points after the endpoint of the transition region belong to place p_2 . The third case corresponds to the same scenario, but with $\tau_w = 3$. In this case, the sensory data associated with the temporal window is considered simply as problematic acquisition. While this data is ignored, the sensory data from the remaining base points is considered as belonging to one place. In the fourth case, there is incoherency at base points x_3 and x_5 . The temporal window starts as $[t_3, t_3]$. While the data associated with x_4 is similar, this does not continue at x_5 . As this is within $\tau_n = 2$, the temporal window extends to $[t_3, t_5]$. As there is no further dissimilarity, the



Fig. 3. Visual features from left to right: Hue, Cartesian (2), hyperbolic (3), intensity.

extent of the temporal window is used to determine the type of incoherency. As it is equal to $\tau_w = 2$, this region is set as transition region and two places are detected. In the last case, the same scenario holds, but this time the extent of the temporal window is not greater than $\tau_w = 2$. This is because $\tau_n = 1$ and there is no dissimilarity in the next frame after x_3 . Therefore, temporal window is not extended. The same holds for dissimilarity at x_5 . Hence, only the dissimilar base points are ignored and there is only one place detected.

V. EXPERIMENTS AND DISCUSSION

This section presents experimental results aimed at evaluating our approach (BuS) in place detection. Two experiments with different datasets are conducted in order to evaluate the performance of our approach. The first dataset is obtained by a Kinect sensor which provides integrated vision and depth data. The second dataset is a vision only dataset which is originally recorded for place categorization purposes. As transition regions are of considerable extent, a transition region is said to be detected correctly if at least 20% is covered within a temporal window.

The selection of the values for informativeness threshold $\tau_{i_1,2}$, coherency threshold τ_3 and two associated coherency parameters τ_n , τ_w and the plenitude threshold τ_p are critical as they directly affect the place detection performance. In these experiments, their values are selected based on manual inspection of the data - except the plenitude threshold that is set as $\tau_p = 1$ in all the experiments. The effect of parameters are further discussed in Sec. V-A. A future work for our current method will be the development of an approach for automatically adapting these parameters based on the incoming data.

A. ImageCLEF 2013 RGB-D Dataset

The first set of experiments are done with an RGB-D dataset recorded for ImageCLEF 2013 Robot Vision challenge. Both the visual and depth data are from a single Kinect frame which has 57.8° horizontal, 43.2° vertical field of view with reliable range measurement interval of $0.6 \leq q_1(b, t) \leq 4$ meters. The first 1,000 frames of the first training sequence is used. Some frames are not informative as shown in Fig. 1 top row. Thus, this experiment allows us to evaluate the effect of depth data on informativeness and coherency checking. Secondly, there are 11 transition regions which corresponds to 12 places that constitute about 84.6% of the data. The feature space \mathcal{L} is of dimension $N_I = 8$. The first feature is depth while the rest are the visual features corresponding to color (hue), Cartesian features, hyperbolic features and intensity as given in Fig. 3. The bubble descriptors are constructed using all the features except intensity. Thus, the length of the vector becomes $N_I = 700$.

The informativeness parameters used during these experiments are given in Table II(a). With these values, about 2.8% of the input sensory data is found to be uninformative. It is observed that having depth data improves performance as the robot is able to determine uninformative frames as shown in Fig. 1. While visual data is found to be informative, this is not the case with depth data.

The variation of performance with respect to different coherency parameter settings are given in Table II(b). When coherency threshold τ_3 is larger, fewer number of base points are considered as incoherent as expected. However, this results in a decrease in correct detection performance. This is probably due to data in the transition regions that are not considered as incoherent so that the actual transition regions are not detected. When the parameter τ_n is varied, incoherent region percentage also varies. This is again expected, since larger τ_n means larger temporal windows and thus, more data resides within these windows. However, a large τ_n can cause a decrease in correct transition detection performance since a large temporal window would contain base points that are actually coherent. As stated before, if more than 80% of the data within a temporal window is coherent then that window is not considered as a correct transition. When the parameter τ_w is larger, less number of temporal windows are formed. As a result, incoherency percentage decreases but the correct detection percentage can also decrease. This is due to the small transition regions with less number of incoherent base points could be ignored when a temporal window is not formed because of size restriction.

Based on the results, the best performance is achieved with $\tau_3 = 10$, $\tau_n = 5$ and $\tau_w = 3$. With these parameters, the incoherency rate of 35.1% is higher than the actual 15.4% which means that certain regions that belong to places have been detected as being incoherent. Such cases occur when sensory data acquisition is problematic such as blurry images or fast rotational camera movements that repeatedly occur in consecutive frames that lead to incoherent interpretation of the data. In this case, an actual single place is detected as

many different places. While there are 11 actual transitions, only 9 of these are detected. Hence, detection rate is about 81.8%.

	τ_{11}	τ_{12}	τ_{81}	τ_{82}
	0.25	0.001	0.2	0.001
Uninformative %	2.8			

(a) Informativeness.

τ_3	10				20			
τ_n	5		10		5		10	
τ_w	3	5	3	5	3	5	3	5
Incoherent %	35.1	33.6	50	49	15.5	13.4	18.9	17.1
Correct Detection %	81.8	72.7	36.4	27.3	45.4	10	45.4	18.2

(b) Place Detection.

TABLE II
IMAGECLEF 2013 DATASET RESULTS.

B. VPC 2009 Dataset

The second dataset is the VPC 2009 dataset that is composed of images from a perspective camera[21]. In the experiments, a subset of data from 3 different homes comprised of 18,300 data points with 40 labeled transition regions is used. The dataset contains uninformative frames as shown in Fig. 1. As two small transition regions from Home 3 sequence are discarded as they occur within the same room, there remains 38 transition regions corresponding to 39 places. Places constitute about 67% of the incoming data sequence on the average while transition regions are associated with data from long (200-300 frames) sequences of base points. The bubble descriptors $I(x, t)$ are constructed based on all the visual features except the intensity and are of dimension $N_I = 600$. The parameters are as given in Table III(a). Differences in environmental and acquisition conditions in regards to each house require using different parameter sets for optimal performance. The experiments are conducted on a Linux PC with 16 GB RAM and Intel Xeon 3.6 Ghz processor. The feature extraction part is developed in C++ while place detection approach is developed in MATLAB.

The results are as presented in Table III(b). From these results, it is observed that the number of detected transition regions vary between 23-57 while the extent of these transition regions vary between 29-76 base points per region. As it can be seen, the feature extraction time of our approach is very competitive compared to other approaches such as SURF and SIFT. Furthermore, the average computation time per transition region varies between 0.003-0.009 with the highest being associated with the third home. This can be attributed to the relatively high number of corresponding base points per transition region in this site.

We also compare the average performance of our approach with other approaches in Table III(c). It is observed that 6.97% of the incoming data is considered uninformative on the average. As transition regions are comprised of long extents of data, these regions themselves are viewed as places. Hence, the average incoherency detection rate is about 32% which is close to the actual value of 33%. The

Home #	τ_{T1}	τ_{T2}	τ_3	τ_n	τ_w
1	0.25	0.005	400	15	10
2	0.2	0.001	15	5	10
3	0.3	0.01	7	15	20

(a) Parameters.

Home #	1	2	3
# Detected TRs	32	57	23
# Base points/TR	37	29	76
Feature Extraction (ms/base point)	300		
Average time per TR (ms/TR)	0.0060	0.0033	0.0094

(b) Place Detection.

Approach		BuS	[16]	[22]
Uninformativeness	%	6.97	NA	NA
Incoherency	%	32	NA	NA
Correct detection	%	84.9	81.6	85.5

(c) Comparative results.

TABLE III
VPC 2009 DATASET RESULTS.

average detection rate of actual transitions is about 84%. Of course, these results depend on the parameter values.

We also compare the average detection rate of our approach with that of [16] and [22] where place detection and recognition are done together with high computational costs due to probabilistic reasoning. With a subset of the VPC 2009 dataset (with 14,346 base points), the detection rates are reported as 81.6% - 85.5%. In our approach, with the data used (18,300 base points) the detection rate is comparable with 84.9%. However, in contrast to these work, our scheme is computationally much simpler since place detection is well separated from place recognition.

VI. CONCLUSION

In this paper, we present a novel approach to place detection in topological maps. Our proposed approach is based on bubble space - where all sensory features and their relative S^2 - geometry are encoded in a manner that is implicitly dependent on robot pose. Its novelty is that ensuring sensory data reliability is integrated with place detection. This is achieved by checking for informativeness, coherence and plenitude using only the bubble space representation of the incoming sensory data. The stringency of these checks is controllable via a set of associated parameters. Our experimental results with two benchmark datasets show that the success rate of the method is comparable with the state-of-the-art approaches. As the robot does not consider unreliable sensory data, it can detect places reliably. Furthermore, the detected places can then be immediately used for node generation in topological maps and for place recognition. As an ongoing work, we are currently studying optimal parameter selection. As future work, we plan to integrate our approach with topological map building in outdoors settings.

ACKNOWLEDGMENT

This work has been supported in part by Tubitak Project EEEAG 111E285, Bogazici University BAP Project 7222

and Turkish State Planning Organization (DPT) under the TAM Project 2007K120610.

REFERENCES

- [1] E. Remolina and B. Kuipers, "Towards a general theory of topological maps," *Artificial Intelligence*, vol. 152, no. 1, pp. 47 – 104, 2004.
- [2] Z. Zivkovic, O. Booij, and B. Kröse, "From images to rooms," *Robotics and Autonomous Systems*, vol. 55, no. 5, pp. 411–418, 2007.
- [3] E. a. Topp and H. I. Christensen, "Detecting Region Transitions for Human-Augmented Mapping," *IEEE Trans. on Rob.*, vol. 26, no. 4, pp. 715–720, 2010.
- [4] P. Beeson, N. Jong, and B. Kuipers, "Towards autonomous topological place detection using the extended voronoi graph," in *Robotics and Automation*, 2005, pp. 4373–4379.
- [5] A. Chella, I. Macaluso, and L. Riano, "Automatic place detection and localization in autonomous robotics," in *IEEE/RSJ Int. Conf. on IROS*, 2007, pp. 741–746.
- [6] A. Tapus and R. Siegwart, "Incremental robot mapping with fingerprints of places," in *IEEE/RSJ Int. Conf. on IROS*, 2005, pp. 2429–2434.
- [7] E. Topp and H. Christensen, "Topological modelling for human augmented mapping," *IEEE/RSJ Int. Conf. on Intell. Rob. and Sys.*, pp. 2257–2263, Oct. 2006.
- [8] H. Korrapati, Y. Mezouar, and P. Martinet, "Efficient Topological Mapping with Image Sequence Partitioning," in *ECMR 2011*, 2011, pp. 1–6.
- [9] Y. Matsumoto, M. Inaba, and H. Inoue, "Visual navigation using view-sequenced route representation," in *IEEE Int. Conf. on Rob. & Aut.*, 1996, pp. 83 – 88.
- [10] F. Fraundorfer, C. Engels, and D. Nister, "Topological mapping, localization and navigation using image collections," in *IEEE/RSJ Int. Conf. on IROS*, 2007, pp. 3872 –3877.
- [11] J. W. Jang and I. K. Oh, "Performance evaluation of scene change detection algorithms," in *Fifth Asia-Pacific Conf. on Communication*, vol. 2, 1999, pp. 841–844 vol.2.
- [12] N. Nourani-Vatani and C. Pradalier, "Scene change detection for vision-based topological mapping and localization," in *IEEE/RSJ Int. Conf. on Intell. Rob. and Sys.*, 2010, pp. 3792–3797.
- [13] J. Martinez-Gomez and B. Caputo, "Towards semi-supervised learning of semantic spatial concepts," in *IEEE Int. Conf. on Rob. and Aut.*, 2011, pp. 1936–1943.
- [14] S. R. Esterby and A. H. El-Shaarawi, "Inference about the point of change in a regression model," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 30, no. 2, pp. 277–285, 1981.
- [15] R. Prescott Adams and D. J. C. MacKay, "Bayesian Online Change-point Detection," *ArXiv e-prints*, Oct. 2007.
- [16] A. Ranganathan, "PLISS: Detecting and labeling places using online change-point detection," in *RSS*, 2010.
- [17] O. Erkent, K. H. and H. I. Bozma, "RGB-D based place representation in topological maps," (to appear) *Machine Vision and Applications*, 2014.
- [18] O. Erkent and H. I. Bozma, "Bubble space and place representation in topological maps," *The Int. J. of Rob. Res.*, vol. 32, no. 6, pp. 671 – 688, 2013.
- [19] G. P. Tolstov, *Fourier Series*. Prentice-Hall, 1962.
- [20] M. Milford, "Vision-based place recognition: how low can you go?" *The Int. J. of Robotics Research*, vol. 32, no. 7, pp. 766–789, 2013.
- [21] J. Wu, H. Christensen, and J. Rehg, "Visual place categorization: Problem, dataset, and algorithm," in *IEEE/RSJ Int. Conf. on IROS*, 2009, pp. 4763–4770.
- [22] A. Ranganathan, "PLISS: labeling places using online changepoint detection," *Autonomous Robots*, vol. 32, no. 4, pp. 351–368, 2012.