Proceedings of the 6th International Conference on Automation, Robotics and Applications, Feb 17-19, 2015, Queenstown, New Zealand

Topological Mapping for Robot Navigation using Affordance Features

Karthik Mahesh Varadarajan
Technical University of Vienna, Austria

***Abstract*— Affordance features are being increasingly used for a number of robotic applications. An open affordance framework called *AfNet* defines over 250 objects in terms of 35 affordance features that are grounded in visual perception algorithms. While *AfNet* is intended for usage with cognitive visual recognition systems, an extension to the framework, called *AfRob* delivers an affordance based ontology targeted at robotic applications. Applications in which *AfRob* has been used include (a) top down task driven saliency detection (b) cognitive object recognition (c) task based object grasping and manipulation. In this paper, we use *AfRob* as base for building topological maps intended for robotic navigation. Traditional approaches to robotic navigation use metric maps or topological maps or hybrid systems that combine the two approaches at different levels of resolution or granularity. While metric and grid based maps provide high accuracy results for optimal path planning schemes, they require high space-time requirements for computation and storage, reducing real-time applicability. On the other hand, topological maps being graph based abstract structures are extremely light and convenient for goal driven navigation, but suffer from lack of resolution, poor self-localization and loop closing. Both approaches show severe restrictions in the case of dynamic environments in which objects which serve as features for the map building procedure are moved or removed from the scene across the time period of usage of the robot. This paper presents a novel approach to topological map building that takes into account affordance features that can help build lightweight, high-resolution, holistic and cognitive maps by predicting positional and functional characteristics of unseen objects. In addition, these features enable a cognitive approach to handling dynamic scene content, providing for enhanced loop closing and self-localization over traditional topological map building. These features also offer cues to place learning and functional room unit classification thereby providing for superior task based path planning. Since these features are easy to detect, fast building of maps is possible. Results on synthetic and real scenes demonstrate the benefits of the proposed approach.**

Keywords-Topology, navigation, metric maps, SLAM, affordances, virtual features, feature prediction, landmarks, place learning, room categorization

I. INTRODUCTION

Conventional algorithms for robot navigation can be classified into three approaches based on types of maps generated. These approaches are (a) metric map based systems (b) topological map based systems and (c) hybrid systems that combine the two approaches at different levels of resolution or granularity [1]. Metric and grid based maps provide high accuracy results for optimal path planning schemes, since they operate on the full pixel or voxel resolution of the perception space of the robot. Hence, they require high space-time requirements for computation and storage, resulting in reduced real-time applicability [2, 3]. Furthermore, since these metric measures are not linked to cognitive or task based symbolic labels, they do not offer the flexibility required for task based or goal based adaptation in robotic path planning. At the other extreme are topological map based approaches that are essentially graph based structures. These structures are extremely light and convenient for goal driven navigation, but suffer from the lack of resolution, since nodes in the graph can be well separated or may not provide the optimal visible path between two waypoints. Furthermore, they are subject to poor self-localization and loop closing due to the lack of discriminability of the features used in such schemes (such as edges) [4,5]. Research in mapping in recent years has been largely geared towards hybrid approaches due to the need to have topological grounding of maps while being able to provide high self-localization and loop closing accuracies [6, 7]. Nevertheless, these approaches only provide for computational trade-offs while being unable to solve key issues in either approach entirely. Recently, semantic information has been used in improving metric maps [12], but lacks scalability. Furthermore, all such approaches show severe restrictions in the case of dynamic environments in which objects which serve as features for the map building procedure are moved or removed from the scene across the time period of usage of the robot. Since the features that are no longer found in the scene of interest lower the recognition rate of the location during a loop closing or localization procedure, dynamic content in the scene as well as objects that can be removed or moved in the scene present serious problems in navigation.

Semantic mapping has also been used in the past for a number of robotic applications. However, explicit topological coupling of semantic maps with a viewpoint to reducing mapping and navigation complexities, while catering to cognitive grounding of the work space model remains unrealized. One such example is Kuiper's classic work on Spatial Semantic Hierarchy [8]. We propose an alternative approach to topological/semantic map building using the concept of affordances or functional object primitives. Affordances have been used in computer vision for object recognition [16, 19, 21, 22, 23, 24, 25], semantic scene processing [14, 15, 17] and robotic manipulation [13, 18, 20]. This paper presents the first attempt to using the concept in the domain of topological and semantic mapping.

II. ALGORITHM

As mentioned above, problems with current state-of-the-art in mapping for navigation is four fold. Firstly, the generated maps can be expensive in terms of estimation time and storage space (as in the case of grid based or pose maps). Secondly, they can have poor self-localization and loop closure (as in the case of topological maps). Thirdly and most importantly (with regard to the main advantage of this paper), none of the schemes can handle dynamic changes in the content of the scene being mapped. Fourthly, cognitive mapping and task based mapping incorporating constraints based on place learning are not efficiently included in the scheme. In our paper, we establish a novel topological mapping procedure that takes advantage of affordance features to overcome these problems. We present features that can provide for high self-localization and loop closure accuracies in a topological mapping framework, thus eliminating the need for metric mapping or grid based local mapping procedures typical of hybrid mapping schemes. Our approach is also cognitively and semantically grounded thereby providing for task based adaptation of path planning algorithms. Hence, we provide a solution that is not just light-weight, high-resolution, holistic and cognitive, but also yields features for place learning and functional room unit classification. These maps are also based on predicted positional and functional characteristics of unseen objects, which is a novel approach to SLAM, essentially making our approach a Virtual Feature SLAM scheme, though our scheme is largely topological and only weakly metric (distance between edges and afforded obstacles are maintained).

The key component in our approach is termed as the affordance feature set. While traditional features for mapping and recognition involve local features such as SIFT or Harris corners or edges, these features are viewpoint dependent. In other words, features representing the same objects can appear different in a variety of poses and environmental conditions. Furthermore, they change with each instance of the object and are not cognitively linked to the functionality of the object. In addition, mapping schemes based on such features do not provide for an elegant mechanism to handle the relocation or removal of such features. While approaches such as RANSAC can handle some of these changes, they do not scale up effectively. In order to alleviate these concerns, we use affordance features. Affordance features are geometric or textural properties of objects that are linked to material and structural characteristics that enable the object to perform a certain function or possess a certain functional property. For example, a cup provides the *contain-ability* affordance that enables it to contain a solid or liquid within its structural bounds. We borrow from [9, 10], the ontology of affordance features for our work in this paper.

A. *Affordance Features, AfNet and Semantic Mapping*

AfNet, The Affordance Network is an open affordance computing initiative that provides affordance knowledge ontologies for common household articles in terms of affordance features using surface forms termed as *afbits* (affordance bits). AfNet currently offers 68 base affordance features (25 structural, 10 material, 33 grasp), providing over 200 object category definitions in terms of 4000 *afbits*. Symbol grounding algorithms for these affordance features enable recognition of objects in visual (RGB-D) data.

Traditional semantic maps for visual scene representation and understanding require the use of information pertaining to at least several hundred objects in the environment of interest. This information – 2D feature and 3D model representations make it impractical for deployment on robots wherein real-time scene processing is pivotal. *AfNet* defines objects in terms of affordance features for efficient semantic storage.

B. *AfRob and Topological Mapping*

The *AfRob* extension [10] to *AfNet* provides linkage for robotic applications. One of the primary schema in *AfRob* is top down search refinement. Since affordances exhibit semantic relationships, this can be taken advantage of in building of topological maps, which is the focus of this paper. *AfRob* defines top down search refinement through semantic affordance features defined by and categorized into three affordance filtration mechanisms. These mechanisms successively reduce the search space of operation in the image to localize and identify objects of interest in a top down search problem. In the approach presented in this paper, we reverse the localization procedure by taking advantage of this positive correlation and establishing hypothetical candidate affordance features in the "local" (current grid location of interest based on view point). These hypotheses are successively refined based on the availability of additional frames of view as the robot explores the scene of interest.

C. *Semantic Affordance Features in AfRob*

The primary features for topological scene processing are the semantic affordance features. Drawing from [10], semantic affordance is dependent on co-occurrence or typical pose of objects. Semantic affordances define the Subject-Object relationships for the functional affordances (structural and material). For example, the semantic affordance corresponding to the co-occurrence of pens and books can be defined based on the engrave-ability functional affordance where the pen and the book form a subject-object pair. In other words, the pen affords engrave-ability on a book. Similarly, the occurrence of a book on a table defines a semantic affordance relationship based on the Support-ability affordance, wherein the table supports the book in this affordance specific pose relationship. Measures of confidence for stand-alone object detection are traditionally augmented by probabilistic reasoning based on detection of other objects in the scene. For example, detection of a table posits the detection of a book on it highly likely, thereby enhancing the confidence levels for detected books.

Thus, semantic affordance features using a concise set of 35 affordance features together with an inference mechanism that reasons the relationships between these affordances yield a system capable of cognitive scene processing that can be extended for topological semantic mapping.

D. *Affordance Filtrations for Topological Mapping*

AfRob formulates the localization of salient regions for visual processing as a three step process. These steps are termed as '*affordance filtrations*' and correspond to identifying entities (affordance grounding mechanisms in actuality) that are related to the queried object through affordance mechanisms. We employ these salient regions for building the fuzzy topological map. For the case of navigation, the localizer affordance filtration creates the primary affordance features, while the secondary affordance features are defined by affordance dual filtrations. The primary and secondary affordances are detected in a multi-resolution framework. In other words, the search for obstacles to navigation proceeds from coarse to fine scales, with granularity being defined by *AfNet*.

i) *Order 0: Localizer Affordances in Obstacle Detection*

The first filtration of the scene for navigation uses the *support-ability* and the *contain-ability* affordance features, along with entity scale information, since these affordance features provide information on the location of object features for structure detection that helps navigation. While the *support-*

ability affordance relates the ability of entities to support objects of a similar or smaller scale, the *contain-ability* affordance relates the ability of entities to enclose or "*contain*" those of a similar or smaller scale. Using the principle that most objects in a given scene are *supported* or *contained* by entities with larger scales helps build global occupancy maps for navigation. In other words, in the given scene of interest, entities that provide the *support-ability* affordance and possess a typical large scale form candidate locations for obstacles. For *e.g*., a table that affords *support-ability* is a candidate for an obstacle in the map. In this case since the *support-ability* affordance is grounded in the abstract geometric concept of flatness/ flat-concavity, detection of all surfaces that satisfy this property enables localization of candidate salient obstacle regions. Similarly, all tables, chairs, desks, shelves in the given scene become candidate obstacle regions, as identified by the affordance, without explicit specification or identification of the objects or entities that provide this affordance, hence bypassing the need for exponentially growing complex semantic linkages and recognition algorithms. In a similar fashion, using the principle that objects can be "*contained*" by entities with larger scales enables detection of obstacles offering this affordance. In other words, entities in the given scene of interest that provide the *contain-ability* affordance also become candidate locations for map boundaries. Based on the scale and the temporal characteristics of the objects, it can be decided if these obstacles relate to static or dynamic obstacles. It should be noted that while static objects are used to define the physical boundaries of the map (*e.g.* large boxes and trash cans), dynamic obstacles are entered into the map with fuzzy confidence measures (per scene observation value of $p_d \sim 0.3$) and can be excluded in the determination of the best fitting position in the map during self-localization and loop closure, since these dynamic objects may have been moved or removed .(*e.g.* yellow blocks *Fig* 2, *row* 5).

ii) Order 1: Affordance Duals and Loop Closing

The second level of affordance filtration for navigation involves identifying all entities (abstracted through affordance grounding mechanisms) that form part of affordance duals. Affordance dualities are affordance feature pairs with opposite subject-object entity relationships. Details are presented in [10]. For *e.g.* *engrave-ability* and *display-ability* form affordance duals since objects capable of engraving can be used on displays or in other words, displays provide a surface for engraving. Thus holes in the topological map can be filled using affordance dual features from visualized affordance features. These hypothetical hole estimates can be refined as new frames become available (for every successive frame i in which d is observed, with p_{d_0} initialized to > 0.5, $p_{d_i} = p_{d_{i-1}} + 1/p_{d_{i-1}}$). These serve to improve the resolution of the topological maps in addition to the localizer affordances that perform hole filling based on unseen objects that are expected to offer *support-ability* and *contain-ability* to seen objects. Duals (affordance grounding structures) that can be found at coarser scales are used to perform the hole filling. For e.g. detection of a structure capable of *engrave-ability* (offered by a 'pen'), can initiate fuzzy map hole filling for candidates that offer *display-ability* (such as 'displays' and 'books'). The detection of objects requiring *contain-ability* affordance expected to be supported in a given position on the map also provides cues to the existence of such entities (that provide *support-ability*) in a position compatible with the object. For example, detection of a large number of structures possessing *contain-ability* affordance predicts a supporting structure below it (and above ground) through duality, the confidence of which can be improved exponentially (bounded by 1, according to equations above) with each iteration of observation (*Fig* 2 *row* 1 for prediction of table structure – red line). Small objects with a *contain-ability* structural affordance, such as cups, mugs, beakers, trays and bowls become dynamic obstacles which are integrated in the map in a fuzzy fashion, with regions supporting them forming static obstacles. The predicted obstacles when observed directly in a different viewpoint enables loop closing, which is a novelty of this approach.

Also, as before, it is not necessary to expressly find all cups, mugs, beakers etc. in the scene – a scheme that is neither scalable not practical. Instead, the direct deployment of the affordance grounding mechanism – detection of the affordance features - in this case, a concavity detector, or a convex superquadric fit, enables determination of candidate locations, thereby bypassing the need for complex object recognition. While detection of an affordance feature dual can aid in hole filling based on the affordance offered by it, the detection of the affordance offered by the object does not necessarily help hole filling using objects possessing the affordance dual. For e.g., as mentioned earlier, detection of displays suggests the presence of *engrave-able* objects such as pens and markers, while the reverse relationship is weakly correlated.

iii) Order 2: Affordance Co-Occurrences for Mapping

The third level of affordance filtration involves identification or localization of entities that provide the same affordance as the queried object but manifest in the scene of interest at larger scales. This is used as a last resort for the topological map filling. For example, if the robot detects a chair like structure in the given scene (that which provides the corresponding affordance), it can be expected that desks and tables can be localized nearby – objects that offer the same affordances – in this case, *support-ability* and *dure-ability* + *dispose-ability* affordances. Hence, slightly different viewpoints containing objects with similar functionality can be mapped to the same location or topological space, since these objects are expected to be found close to each other, though there may not be any common objects observed in both scenes.

It should be noted that the level of importance of the various filtrations proceeds in the order listed. In other words, the first level of filtration based on the *support-ability* and *contain-ability* affordances form the primal filtration mechanism for mapping and in this scenario is purely uni-directional and at the coarsest scale. The second level of filtration forms the secondary level – prediction through duality and the third or tertiary level of filtration aids the primary mechanism through associative mapping. Examples of filtrations are presented from [10].

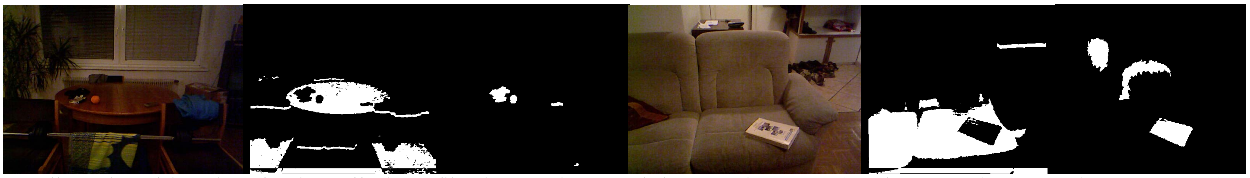
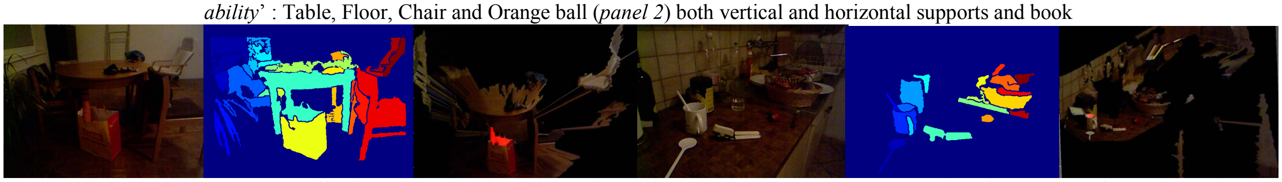
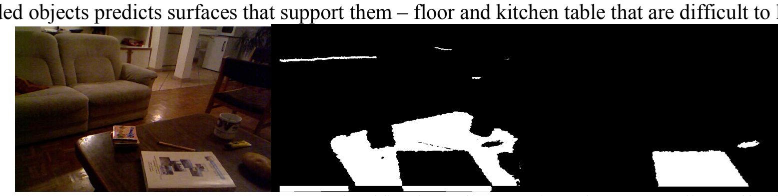


Figure 1. Illustration of detection of various affordance filtration cues in a sample home environment (a) Order 0: *Localizer + Dual augmentation (panel 1)* Affordance filtration through localizer affordance- *support-ability* – detection of stable supports augmented by affordance dual objects providing '*Contain-ability*' : Table, Floor, Chair and Orange ball (*panel 2*) both vertical and horizontal supports and book



(b). Order 1: *Dual based Prediction* - Affordance filtration through localizer affordance: *Contain-ability* – Detection of concave surfaces (shown in column 3) localizing afforded objects predicts surfaces that support them – floor and kitchen table that are difficult to localize due to clutter



(c) Order 2: *Localizer + Dual and Co-occurrence based Prediction*: Identification of supporting planes as well as affordance filtration through affordance duals - Estimation of book regions from locality for detected pens: *Engrave-ability* dual. These objects predict and augment confidence of the supporting table which is an obstacle.

E. Topological Mapping based on AfRob features

While it is possible to use point-cloud based 3D metric schemes augmenting Laser data to build spatial maps, such high resolutions are unnecessary (and yet less accurate in terms of semantic mapping) from the standpoint of robot navigation. Such 3D based affordance gridding mechanisms are well suited for full 6 DOF localization and 3D map building. In the schema presented in this paper, we restrict our map building process to a graph based 2D projection that holds boundary/ obstacle information based on high confidence measures from the primary affordances and accumulative probabilistic measures from the secondary affordances at each point on the map. Specifically, as shown in *Fig* 2, our topological map is a metric graph $G_t(V_t, E_t)$, embedded in an occupancy grid O, with nodes being added based on observed bounds from laser data, afforded objects or obstacles, predicted objects based on duality/ co-occurrences, robot locations ($V_t = V_r \cup V_{ao} \cup V_{ap} \cup V_{ac}$) and augmented with distance information represented in the form of edge weights ($E_t = \forall_o E_d \cup \forall_b E_l$). The edge weights are obtained using depth distances for all obstacles as well as laser data for all large bounds. The implicit edge linkage angles ($E_t(\theta_{ij}) = \angle(E_i, E_j) = |r_{\theta i} - r_{\theta j}|$) have been obtained using robot odometry (or using depth scene centroid). Exploration for mapping can be done manually or using scene change estimation, which involves addition of nodes to the graph when new affordance entities are observed. Probabilities of dynamic and predicted objects are updated based on repeated predictive observation as well as direct observation (in which case the $p_d = 1$). As described earlier, we use affordances as feature descriptors in contrast to edges used in topological maps: hence providing better feature resolution as well as uniqueness and discernibility in addition to grounding the actual functional object in the scene (intrinsically tied to object recognition). This approach is not as expensive as metric maps, but also provides far better accuracy than topological maps, since it provides for predictive loop closing (and metric maps under dynamic conditions). Since metric maps use metric features of objects, they are not flexible to changes and result in over-fitting errors. The map building procedure is summarized in *Algorithm* 1. The primary affordance mechanisms provide evidence for strong feature hypotheses, while the secondary affordance mechanisms provide evidence towards weak feature hypotheses. The strong feature hypotheses pertain to objects that are essential to complete the topological map and can be assumed to exist in the given scene of interest, despite the lack of visible presence in the field of view of the camera. On the other hand, the weak feature hypotheses or fuzzy feature hypotheses pertain to objects that might possibly exist in the given scene, but the presence or location of which cannot be determined with certainty from the given field of view. Evidence from successive image frames or frames from alternate viewpoints is necessary to establish the presence of such objects. Following the algorithm, the generation of topological maps using primary and secondary affordance features, laser data (used to fit large linear structures in the occupancy map) and robot odometry (for angle between edges in graph) for a simulation environment and two domestic environments is shown in *fig* 2 and 3, 4 respectively. For the sake of ease of representation the output graph is made to contain only robot position nodes and the output occupancy grid contains filled locations from the graph node positions that don't correspond to robot positions.

Algorithm *1. Topological Map Generation*

Initialize $G_t(V_t, E_t) = \emptyset$, $\forall_{i,j} \mathrm{O}(i,j) = \emptyset$
Iterate over viewpoint k,
  Identify all v_{ao}, v_{ap}, v_{ac} (elements of V_{ao}, V_{ap}, V_{ac})
  If any $v_{ao} \sim v_{ap_m}$, where $m = \{1..k-1\}$,
  $p_{ap_m} = 1, c_{ap_m} = c_{ao}$: *loop closing*
  If any $v_{ap} \sim v_{ap_m}$, where $m = \{1..k-1\}$,
  $p_{ap} = p_{ap_{k-1}} + \frac{1}{p_{ap_{k-1}}}, c_{ap} = mean(c_{ap}, c_{ap_{k-1}})$:

For all other v, set $p_{ao} = 1, p_{ap} = 0.7, p_{ac} = 0.65$
Estimate E_d for all obstacles and E_l for all bounds
$E_t(\theta_{ij}) = \angle(E_i, E_j) = |r_{\theta i} - r_{\theta j}|), \forall\, r_{ij}$
$E_t(\theta_{ij}) = \angle(E_i, E_j) = |r_{\theta i} - c_{\theta j}|), \forall\, r_i, c_j = \{o, b\}_j$
$G_t(V_t, E_t) = G_t(V_t, E_t) \cup \{V_r \cup V_{ao} \cup V_{ap} \cup V_{ac}, \forall_o E_d \cup \forall_b E_l\}$
Output $O(x, y) = 1\ \forall_{t \neq r}\ t_{x,y}, G_t(V_t, E_t) = G_t(V_t, E_t) \forall_{t=r}$

While detection of entities providing the *support-ability* affordance, such as a table, a bed, a cupboard or a shelf relate to structure in the scene which have a lower incidence of being moved, detection of entities providing secondary affordance mechanisms such as *engrave-ability*, which is afforded by a pen, is tied to the high incidence of relocation of the entity. This information is incorporated into the estimation of the topological map and serves as highly useful features that overcome the errors arising from the rigid nature of traditional metric maps. Since metric maps use exact object features for remapping and merging maps, they produce erroneous results in the cases of objects that have been moved or removed from the environment. Approaches such as *RANSAC* are used to provide the best estimate matching to enable map merging. However, since such approaches do not distinguish between scene structure and movable entities in the scene of interest, they produce poor results. Our approach can ascertain whether an object is part of the static structure of the scene or is capable of being moved from a single frame, in a probabilistic manner and without the sensor actually requiring the movement of the object to be observed during the building of the map.

F. Place Learning and Map Enhancement

Affordance features can also be used in the high level topological mapping process. Specifics of determining boundary estimates for mapping based on walls and doorways that we user here are from [11]. Features such as number of doors, number and size of walls etc. have been used for detection of functionality of rooms and place learning. The use of affordance features enables improvement over this scheme. For example, in a domestic environment, the detection of a significant number of entities providing the *support-ability* affordance at the height of typical tables, chairs, couches, desks and cupboards provides probabilistic evidence towards the categorization of the room as a hall or drawing room. On the other hand, detection of a few large surfaces possessing the *support-ability* affordance and at the height of a typical bed provides evidence towards the functional categorization of the room as a bedroom. Detection of a large number of objects that provide the *contain-ability* affordance such as plates, bowls, mugs etc. and that which provide the *intrinsic contain-ability* affordance such as canisters, food packs and cartons, tetrapaks etc. help categorize the room as a kitchen or cooking space. Similar affordance feature based evidentiary reasoning and functional categorization can be extended to other rooms in a domestic environment or in other work spaces of deployment of the robot. The use of affordance features results in accurate room functional categorization for scenes in *Fig* 2, 3, 4 (4/4, 4/4 and 2/2 as opposed to 2/4, 2/4, 2/2 for [11]). It should be noted that since cognitive features such as affordance features are used in the schema for topological mapping, the detection of room functionality is independent of the exact objects in the scene. In other words, while traditional functionality detection algorithms depend on pre-learnt features such as point features or texture patterns on mugs and bowls in a kitchen or desks and couch in a drawing room, these could change arbitrary in typical work environment, wherein the location or existence of such entities might vary with time. On the other hand, room functionality detection using affordance features is robust to such changes. Furthermore, this information can be incorporated into the topological mapping process. Topological mapping with higher level functional categorization or place learning enables better cognitive mapping of the work space of the robot and aids the robot in performing optimal decisions in terms of navigation for tasks execution. For example, if a robot was to use such topological maps for its path planning with the robot being present in the hall and requiring to bring a mug from the kitchen, the robot can directly ascertain the shortest navigation route to reach the kitchen and use this information to bring the mug to the user in the shortest possible time and with minimal effort.

III. Integrated Affordance Topology Maps

The results of generation of affordance filtration based topological maps in indoor environments is demonstrated through a simulation scenario described in *Fig* 2. In *row* 1 of the image, the objects providing the *contain-ability* affordance (denoted in red outlines) and ascertained using algorithm from [9, 10] (working on depth images), predict the presence of a supportable surface (marked by a red line) that is added to the map (with $p \sim 0.7$), even though this is not observed directly. As the robot returns to the kitchen table, as demonstrated in row 4, it sees a larger section of the kitchen table and the predicted supportable surface added to the map earlier with a low confidence value is now confirmed and augmented with a higher confidence value. In essence, this enables probabilistic loop closing. The presence of a large number of containers also indicate that the robot is present in a kitchen like environment. The supports that are observed directly are marked using red circles in the images, while bounds of environment, such as walls are marked using violet circles. In *row* 2, the robot observes the bedroom and identifies the low supporting surfaces and wall bounds and uses this information to arrive at the conclusion that the likely environment is a bedroom. As before, obstacles are added to the graph. In *row* 3, the robot moves towards the bathroom wherein the presence of few objects providing *contain-ability* indicate the functionality of the room. When the robot returns to the kitchen as denoted in *row* 4, it observes the supporting surface of the table and corrects the position and confidence of the obstacle. In *row* 5, the robot has moved towards the hall, where a dynamic obstacle (indicated by yellow outlines) is observed. This is placed in the map as a transient obstacle but is not used for further matching and localization as it is not observed in successive frames. In *row* 6, the robot observes the supporting surfaces (with given height and size) from the tables and couches indicating that the robot is in the hall. As can be seen, the size of the graph (6 nodes) and complexity of adding new transient nodes (~3 per waypoint) is quite low. Also notable is efficient loop closing or node merging. Error at loop closing (being a simulated environment) with accurate range measurement was ~0. With a traditional topological map the obstacle marked in red in *row* 1 is not identified resulting in possible collisions with the obstacle. Note that weights have been omitted in the final map for clarity. The results of the mapping procedure in practical domestic scenarios were also evaluated. The results of the affordance filtration based feature detection and topological mapping is presented for two domestic environment scenarios in *Fig* 3 and 4. The corresponding topological maps generated and the results of functional classification of various rooms in the environment of interest are also shown step by step. It can be seen that using such an approach for affordance based grouping and determination of occupancy of regions results in a suitable graphical topological path through the scene of interest. The blocks in the topological maps shown in *Fig* 3 and 4, in different colors indicate different types of affordance feature clusters

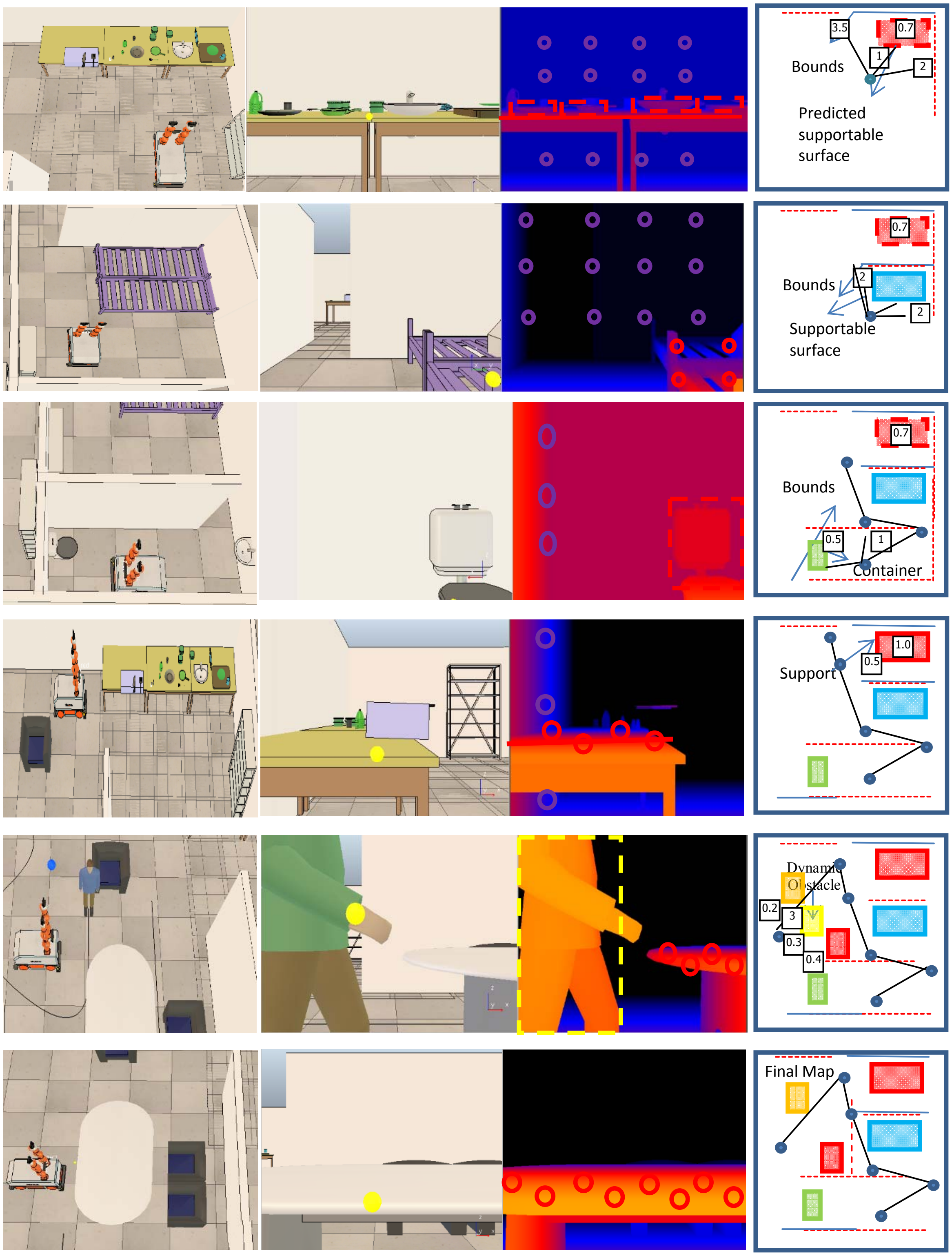
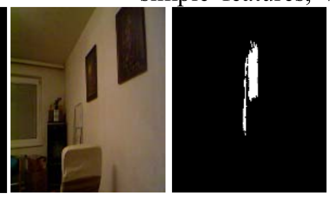


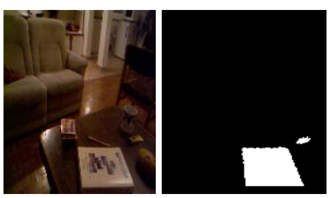
Figure 2. Affordance based topology detection in simulation

detected. Red indicates supportable surfaces such as kitchen tables. Blue corresponds to surfaces similar to couches/chairs and yellow to surfaces of size and height similar to that of a bed. This functional categorization of the room can help in task based navigation. Number of nodes in the graph for the two scenarios were 8, 4 respectively and error in position at loop closing (node merging) less than 0.25m.

Support-ability cues

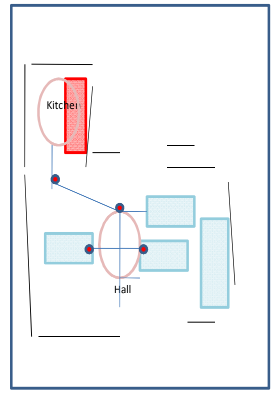


Display-ability cues



Engrave-ability cues

Contain-ability cues



Final map

Figure 3. Topological map building in domestic environment – scenario 1

IV. Conclusion

In this paper, we have demonstrated a novel approach towards using affordance features for navigation by building topological maps that are based on predictive functional obstacles. We have also shown through simulated scenes and controlled topological map building scenarios that the usage of predicted affordances improves over traditional topological feature maps. Furthermore, the features used in the affordance estimation are simple. While using such simple features, we enable better self-localization and loop closing, especially in dynamic scenarios. Future work will involve development of affordance derived metric maps to enable full functional SLAM integration.

V. References

[1] S. Thrun, "Learning metric-topological maps for indoor mobile robot navigation," Artificial Intelligence, vol. 99, no. 1, pp. 21–71, 1998.

[2] C. Estrada, J. Neira, and J. Tardos, "Hierarchical slam: Real-time accurate mapping of large environments," IEEE Transactions on Robotics, vol. 21, no. 4, 2005.

[3] S. Thrun and M. Montemerlo, "The GraphSLAM algorithm with applications to large-scale mapping of urban structures," International Journal on Robotics Research, vol. 25, no. 5/6, pp. 403–430, 2005.

[4] E. Fabrizi and A. Saffiotti, "Augmenting topology-based maps with geometric information," RAS, vol. 40, no. 2, 2002.

[5] D. Rawlinson and R. Jarvis, "Topologically-directed navigation," Robotica, vol. 26, no. 2, pp. 189–203, 2008

[6] Navigation in Hybrid Metric–Topological Maps, Kurt Konolige, Eitan Marder-Eppstein, Bhaskara Marthi, ICRA 2011.

[7] K. Ni, D. Steedly, and F. Dellaert, "Tectonic slam: exact, out-of-core, submap-based slam," in In Proc. IEEE International Conference on Robotics and Automation, 2007, pp. 1678–1685..

[8] B. Kuipers and Y.-T. Byun, "A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations," Journal of Robotics and Autonomous Systems, vol. 8, pp. 47–63, 1991.

[9] KM Varadarajan, M. Vincze, "AfNet: The affordance network", IEEE Asian Conference on Computer Vision (ACCV) 2012.

[10] KM Varadarajan, M. Vincze, "AfRob: Affordance Network for Robotics", International Conference on Robotics and Systems, 2012.

[11] KM. Varadarajan, M. Vincze, "3D room modeling and doorway detection from indoor stereo imagery using feature guided piecewise depth diffusion", International Conference on Robotics and Systems, 2010.

[12] MR. Walter, S. Hemachandra, B. Homberg, S. Tellex, and S. Teller "Learning Semantic Maps from Natural Language Descriptions", Robotics Science and Systems, 2013.

[13] Koppula, Hema S., and Ashutosh Saxena. "Anticipating human activities using object affordances for reactive robotic response." RSS, 2013.

[14] Jiang, Yun, and Ashutosh Saxena. "Infinite latent conditional random fields for modeling environments through humans." RSS, 2013.

[15] Jiang, Yun, Hema Koppula, and Ashutosh Saxena. "Hallucinated humans as the hidden context for labeling 3d scenes." CVPR, 2013.

[16] Yao, Bangpeng, Jiayuan Ma, and Li Fei-Fei. "Discovering object functionality." IEEE International Conference on Computer Vision 2013.

[17] Chu, Vivian, et al. "Using Robotic Exploratory Procedures to Learn the Meaning of Haptic Adjectives", ICRA 2013.

[18] Song, Hyun Oh, et al. "Visual grasp affordances from appearance-based cues." IEEE Computer Vision Workshops (ICCV Workshops), 2011

[19] Bar-Aviv, Ezer, and Ehud Rivlin. "Functional 3D Object Classification Using Simulation of Embodied Agent." BMVC. 2006.

[20] Zhao, Yibiao, and Song-Chun Zhu. "Scene Parsing by Integrating Function, Geometry and Appearance Models." CVPR, 2013.

[21] Grabner, Helmut, Juergen Gall, and Luc Van Gool. "What makes a chair a chair?." Computer Vision and Pattern Recognition (CVPR), 2011

[22] Gall, Juergen, Andrea Fossati, and Luc Van Gool. "Functional categorization of objects using real-time markerless motion capture." Computer Vision and Pattern Recognition (CVPR), 2011.

[23] B Tversky, K Hemenway, 'Form and Function', Journal of Experimental Psychology, 2001.

[24] B Tversky, K Hemenway, 'Objects, Parts and Categories', Journal of Experimental Psychology, 1997.

[25] Rivlin, Ehud, Sven J. Dickinson, and Azriel Rosenfeld. "Recognition by functional parts [function-based object recognition]." CVPR 1994.

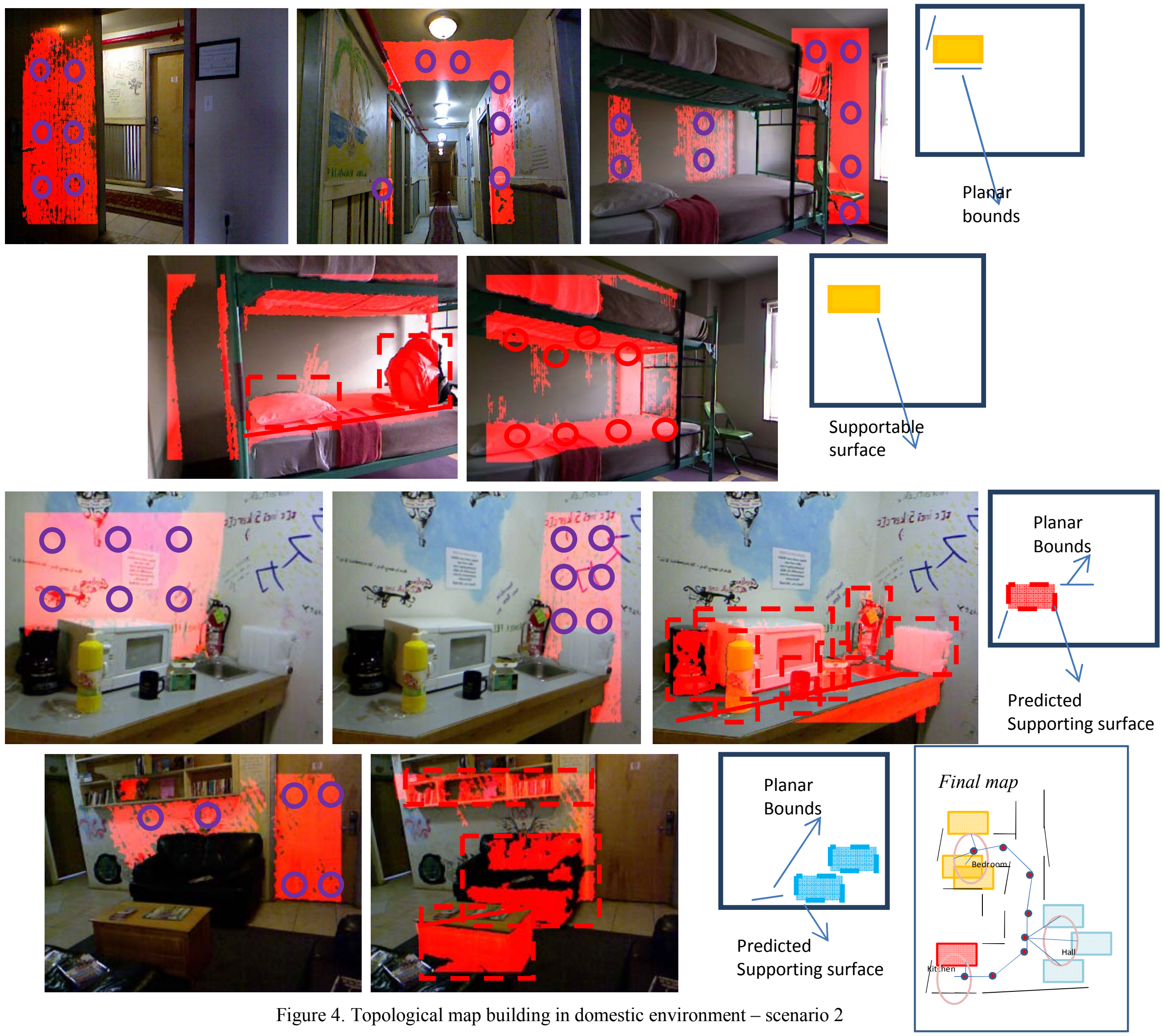


Figure 4. Topological map building in domestic environment – scenario 2