FAB-MAP 3D: Topological Mapping with Spatial and Visual Appearance

Rohan Paul and Paul Newman
Oxford University Mobile Robotics Research Group. {rohanp,pnewman}@robots.ox.ac.uk

***Abstract*—We present a probabilistic framework for appearance based navigation and mapping using spatial and visual appearance data. We adopt a bag-of-words approach in which positive or negative observations of visual words in a scene are used to discriminate between already visited and new places. Additionally, we explicitly model the spatial distribution of visual words as a 3D random graph in which nodes are visual words and edges are distributions over distances. The spatial model captures the multi-modal distributions of inter-word spacing and incorporates a probabilistic sensor model for word detection and distances. Crucially, the inter-word distances in 3D are viewpoint invariant and collectively constitute strong place signatures for appearance based navigation. Results illustrate a tremendous increase in precision-recall area compared to a state-of-the-art visual appearance only systems.**

The goal of this work is non-metric topological navigation and mapping in appearance space - a by-product of which is loop closure detection. We provide and test a formulation which uses not only the visual appearance of scenes but also aspects of its geometry. Our approach, called FAB-MAP 3D, has its roots in the FAB-MAP algorithm which in essence learns a probabilistic model of scene appearance online using a generative model of visual word observations and a sensor model which explains missed observations of visual words. FAB-MAP 3D takes the same approach but incorporates the observation of spatial ranges between words coupled to the observation of pairs of visual words, Figure 1. This interaction is captured via a random graph which models a distribution over word occurrences as well as their pairwise distances. Using non-parametric Kernel Density Estimation we learn complex multi-modal distributions over inter-word distances and also accelerate inference by executing a Delaunay tessellation of the perceived 3D graph. The system shows improved performance over vision only sensing in an outdoor setting.

Our motivation for incorporating range information is two fold. Firstly, prior to this the work, the FAB-MAP framework only modeled the presence or absence of a word at a location and did not incorporate the spatial arrangement of visual words. Secondly, FAB-MAP currently discards the number of times a word appears in a scene - there is information being neglected here. This is addressed in FAB-MAP 3D because by using the range between occurrences of visual words we are implicitly counting word occurrence. Note also that we are in the business of robotics where range information is ubiquitous be it from lidar, stereo or structure from motion - we should use it if we can. Finally, there is also an important prima facia advantage of using distances because they are invariant under rigid transformation and that is precisely what we require of a place descriptor in topological navigation. The system only needs intra-scene distances which can be derived in a local frame without requiring a global metric map.

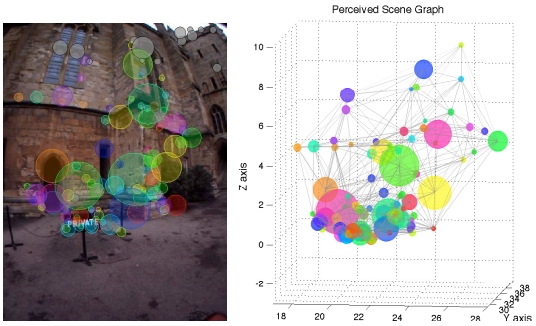

Figure 1. Random graph location model. Robot view (left) and perceived 3D constellation of visual features (right). 3D distances from lidar, stereo or structure from motion.

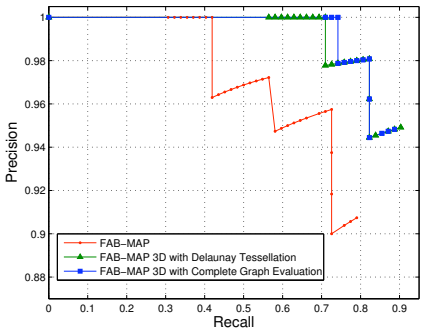

Figure 2. Precision-recall curves for the New College data set. FAB-MAP 3D has a higher recall of 74% at 100% than FAB-MAP that has 42% recall at 100% precision. The accelerated approach has marginally lower recall of 71% but still performs better than FAB-MAP.

FAB-MAP 3D provides substantial and compelling improvement in precision-recall performance over the existing FAB-MAP system, Figure 2. By capturing spatial information, the algorithm reduces the number of false positives and shows a dramatic decrease in false negative rate, particularly in scenes possessing a large number of common words where a loop closure decision hinges on spatial information. The framework shows robustness to perceptual aliasing as well as scene change. The system scales linearly with the number of places in the map. Graph inference can be accelerated by executing a Delaunay tessellation of the observed graph scaling log-linearly with scene complexity.