Global Localization in a Dense Continuous Topological Map

Edward Johns and Guang-Zhong Yang
The Hamlyn Centre, Imperial College London
{ej09, g.z.yang}@imperial.ac.uk

***Abstract*—Vision-based topological maps for mobile robot localization traditionally consist of a set of images captured along a path, with a query image then compared to every individual map image. This paper introduces a new approach to topological mapping, whereby the map consists of a set of landmarks that are detected across multiple images, spanning the continuous space between nodal images. Matches are then made to landmarks, rather than to individual images, enabling a topological map of far greater density than traditionally possible, without sacrificing computational speed. Furthermore, by treating each landmark independently, a probabilistic approach to localization can be employed by taking into account the learned discriminative properties of each landmark. An optimization stage is then used to adjust the map according to speed and localization accuracy requirements. Results for global localization show a greater positive location identification rate compared to the traditional topological map, together with enabling a greater localization resolution in the denser topological map, without requiring a decrease in frame rate.**

I. Introduction

Vision-based approaches to navigation are popular within the mobile robotics community due to the low-cost of vision sensors, the large quantity of data gathered, and the close relationship to the human sensory system. Visual localization, as a major component of an overall navigation system, has been the subject of intense research attention for many years, but there remain significant challenges. These include dealing with dynamic environments (changes in illumination, long-term changes in environment structure), mapping strategies (robustness to localization errors, online map building), perceptual aliasing (similar observations at different locations), and computational issues (processing speed, memory requirements). Approaches to localization can generally be categorized as either metric [1][2], where the pose of the robot is estimated within a defined coordinate system, or topological [3][4], where a query image is compared to other images previously captured at discrete locations within the environment, often along a fixed path that the robot is expected to follow.

If the initial location of a robot is known, then probabilistic localization can be carried out by use of the Kalman filter [5]. However, if the initial location is unknown, or to deal with re-localization after the "kidnapped robot" problem, a probability distribution can be computed across all possible locations, known as *global localization*. For a sensor reading z, this requires the computation of $p(z \mid x_k)$ across all possible states $x_{1 \ldots k}$. With a topological approach, z_q represents a query image from the robot's camera, and x_k represents the location, or node, of the robot within the topology. $p(z_q \mid x_k)$ is then computed via a similarity measurement between image z_q and the image z_k which was captured during a prior map-building stage. The location probability distribution can then be updated from the currently predicted location distribution, with techniques such as particle filters [6].

As with traditional topological localization methods, this paper presents a technique that calculates the similarity between a captured image and those in the topological map, in order to compute $p(z_q \mid x_k)$. However, one of the issues with traditional topological approaches is that if nodes are inserted at too great a separation, then the localization resolution may suffer. Methods exist to estimate the robot's pose relative to sparsely-spaced nodes [19], but such methods are noisy compared to maps with a higher density of nodes. Inserting nodes at a greater spatial frequency can improve localization accuracy, but at a significant cost to computational speed, as every image in the map must be individually compared to the query image.

The work described in this paper presents a new approach to topological mapping which allows for a very dense topology without sacrificing computational speed. From a sequence of images acquired by a robot, local features are tracked across multiple adjacent images to create a set of landmarks representing real-world points, spanning the continuous topological space between each nodal image in the map. During localization, features from a query image are matched to these landmarks, rather than to the individual images as in the traditional approaches. Inserting further nodes into the topological map does not significantly increase the number of landmarks, as the landmarks often already exist from detection in the adjacent nodes. In this way, the topological density can be very high without inducing the penalty of a large decrease in frame rate. Fig. 1 demonstrates the key difference between the proposed approach and the traditional approach.

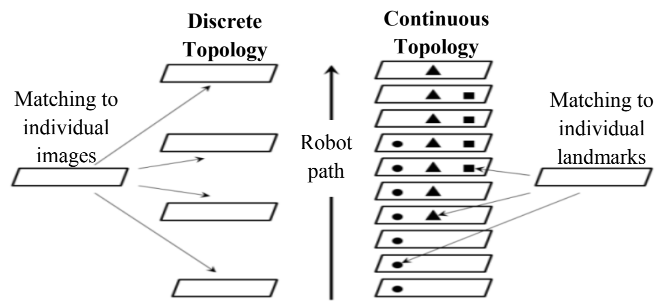

Fig. 1. Comparison of the traditional discrete topological map to the dense continuous topological map introduced in this paper.

Furthermore, by observing features from a number of different viewpoints, an understanding is learned of how each feature is expected to change as the robot traverses the environment. This allows a probabilistic approach to the image similarity measure, rather than the more naive voting scheme that is often adopted. Finally, learning each landmark's independent behaviour also enables optimization of the map with respect to the scale of the environment and the required accuracy-speed trade-off.

II. RELATED WORK

The task of finding the similarity between an image from the robot's current view and all nodal images in the topological map, draws from standard image matching techniques in wider computer vision literature [7][8]. Images can be represented in a holistic manner by considering the image in its entirety [9], by extracting local features [10] detected at interest points, or by combining both methods in a hybrid approach [11]. Whilst matching many local features is more time consuming than matching a single global feature, it offers far greater robustness to occlusions and illuminations effects, as well as offering greater discriminative power. Several local features exist [12] which demonstrate robustness to scale, rotation and illumination, and small affine viewpoint changes. The method presented in this paper uses SIFT features [13], which describe the texture around a keypoint using histograms of gradient orientations, although the same methodology works for any local feature that offers some tolerance to affine viewpoint changes.

Matching features between two images often involves computing the distance between feature descriptors, and classifying a match when the distance between the two closest feature descriptors is less than the second smallest distance, by some threshold [13]. Geometric constraints have also been developed to prune out false matches [14] by considering the spatial arrangements of nearby features. Whilst this simple technique can perform well, as the size of the image database increases, the number of matches that pass this threshold decreases, as each feature becomes less discriminative within the larger pool. In the method proposed in this paper, by tracking features across images and learning their expected descriptor variation, every feature can be matched.

Topological approaches to localization use image sequences to represent a map, with matching performed between a query image and those images in the map, using one of the many image matching techniques discussed. Those employing local features have had success with simple feature voting schemes [10]. Tracking features across multiple images to generate landmarks, similar to the proposal in this paper, has also been demonstrated as an effective means of retaining only the most stable features [16][18]. Computing the distinctiveness of local features, again investigated in this paper, has shown to provide more efficient image matching using a probabilistic approach [17]. This paper draws from the work on feature tracking and probabilistic matching to propose a new technique for dealing with topological localization with a far greater nodal density.

III. GENERATING LANDMARKS

Generation of the landmarks for the map requires a mobile robot to travel along a path and capture images sequentially. SIFT features [13] are detected in each image, and then tracked through adjacent images by finding corresponding feature matches. A potential match is found if the feature descriptor distance between the two features is less than a threshold. By requiring less than 1% false feature matches, the threshold was empirically set at 0.45 for the standard normalized 128-dimensinal SIFT descriptor. False matches are then pruned out by estimating the epipolar geometry between the two adjacent images using the RANSAC method [15]. A feature which is matched is then tracked continually across each consecutive image, until a match is no longer found.

From each feature track, a landmark is then generated and described by a set of properties, computed from the features present in the feature track. The mean descriptor, d_{mean}, is computed across all features in the track, together with the maximum descriptor distance, d_{max}, between every feature in the track and the mean descriptor. With each landmark then occupying a volume in feature space, rather than a single point, it is expected that a large number of false positive matches may occur between a query feature and a landmark. As such, further discrimination is added to each landmark by including an additional neighbour landmark for geometric verification of a match. Thus, for each landmark, every other landmark that co-occurs in at least one image is assigned as a potential neighbour, and the minimum and maximum spatial distance, δ_{min} and δ_{max}, and angle, θ_{min} and θ_{max}, are calculated across all co-occurring images. Finally, every image captured along the robot's path, representing a node in the topology, is assigned a set of landmarks which appear in the image. In summary thus far, the map is represented by a set of nodes, or images, $\mathbf{x}$, with each image $x \in \mathbf{x}$ represented by a set of landmarks $\mathbf{l}$, where every landmark $l \in \mathbf{l}$ is described by $l = \{d_{mean}, d_{max}, \mathbf{l_n}\}$, and every neighbour landmark $l_n \in \mathbf{l_n}$ is described by $l_n = \{\delta_{min}, \delta_{max}, \theta_{min}, \theta_{max}\}$.

Fig. 2 shows an example of two landmarks that have been generated by three features each. The landmark representing the light on the ceiling will be represented by, of the two, a smaller descriptor range, due to the lack of background clutter, and also the smaller angular viewpoint change across the features. As will be detailed later, this landmark will therefore have a greater weighting in the overall image similarity calculation to take advantage of its greater discriminative power.

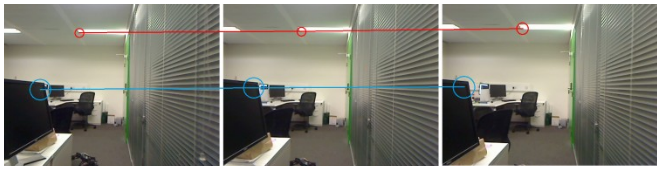

Fig. 2. Landmarks detected across multiple viewpoints. The landmark on the ceiling has a smaller descriptor range across all features in the track, due to its greater distance from the camera, together with its lack of background clutter. This landmark therefore has a greater weighting in the image similarity measure.

IV. MATCHING FEATURES TO LANDMARKS

With the map having been generated and the robot now in the localization phase, features are extracted from a query image and a match is attempted between each feature f in the image, and each landmark l from the map. For each feature, the descriptor d_f is computed, together with the descriptor, d_n, of every neighbouring feature, f_n, in the same image. The spatial distance δ_n and angle θ_n between feature f and neighbouring feature f_n are also computed. Fig. 3 shows the Bayesian network used to compute the probability of a landmark match given these measurements from a query image.

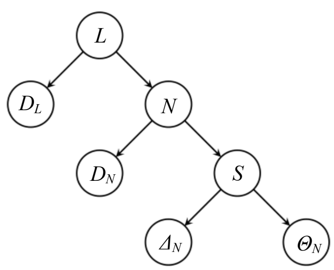

Fig. 3. Bayesian network used to compute the probability of a landmark match, given observations from a query image.

In the Bayesian network of Fig. 3, L, N and S represent beliefs about the presence of a landmark, l, the presence of a neighbour landmark, l_n, and the spatial arrangement of the two, respectively, for every attempted match with a query feature f. L represents the belief that feature f is a true positive match to landmark l, and N represents the belief that there exists a neighbour feature f_n, out of all possible neighbour features, which is a true positive match to neighbour landmark l_n. S then represents the belief that f and f_n lie within the expected spatial limits.

D_L, D_N, S, Δ_N and Θ_N represent the actual measurements to be taken from features f and f_n. D_L indicates whether the descriptor, d_f, of feature f, lies within the descriptor range of landmark l. As previously discussed, a landmark is described by $l = \{d_{mean}, d_{max}, \mathbf{l_n}\}$. Thus, D_L is equal to 1 *iff* $|d_f - d_{mean}| \le d_{\max}$, and 0 otherwise. In a similar manner, D_N indicates whether the descriptor, d_n, of a neighbour feature f_n, lies within the descriptor range of the neighbour landmark l_n. Θ_N and Δ_N indicate whether the neighbour feature lies within the correct angular and distance ranges, respectively. As before, a neighbour landmark is described by $l_n = \{\delta_{min}, \delta_{max}, \theta_{min}, \theta_{max}\}$. Thus, Δ_N is equal to 1 *iff* $\delta_n \ge \delta_{\min}$ *and* $\delta_n \le \delta_{\max}$, and 0 otherwise. Similarly, Θ_N is equal to 1 *iff* $\theta_n \ge \theta_{\min}$ *and* $\theta_n \le \theta_{\max}$, and 0 otherwise. Finally, S, is the belief that the neighbour feature f_n lies within the correct overall spatial arrangement relative to feature f, is equal to 1 *iff* $\Theta_N = 1$ *and* $\Delta_N = 1$, and 0 otherwise.

The Bayesian network described provides the core to the probabilistic landmark matching. For a query feature f giving observations $D_L = 1$, $D_N = 1$, $\Delta_N = 1$ and $\Theta_N = 1$, the probability of a true positive match to a landmark l can then be calculated with Equation (1). If any of the measurements are not equal to 1, then the probability of a match is zero. Equation (1) is computed for every query feature f, and the maximum value across all query features is taken as the probability that there exists, in the query image, a true positive match to landmark l.

$$\begin{aligned} & p(L=1 \mid D_L=1, D_N=1, S_N=1) = p(L=1 \mid E_L=1) \\ & = \frac{p(L=1)\,p(E_L \mid L=1)}{\sum_{L=0,1} p(E_L=1 \mid L)\,p(L)} \end{aligned} \tag{1}$$

Here, E_L represents the combined evidence of D_L, D_N, and S_N, and is equal to 1 *iff* these three measurements are equal to 1. The denominator in Equation (1) represents the overall value of $p(D_L = 1, D_N = 1, S = 1)$ and is marginalized over L, to include the possibility of false positive matches from the given observations. For global localization, $p(L) = a/b$, where a is the number of features in the feature track from which landmark l was originally built, and b is the total number of features detected in the original map-building tour. The value of $p(E_L = 1 \mid L = 1)$ is 1 because the ranges of D_L, D_N and S were originally computed directly from true landmarks. However, the probability of these observations given that, in fact, the landmark is not present, requires marginalizing over the states of N, as in Equation (2).

$$\begin{aligned} & p(E_L = 1 \mid L = 0) = \\ & p(D_L = 1 \mid L = 0) \sum_{N=0,1} p(D_N = 1 \mid N)\,p(S = 1 \mid N)\,p(N \mid L = 0) \end{aligned} \tag{2}$$

Computing the probability $p(D_L = 1 \mid L = 0)$ of a descriptor match between feature f and landmark l, given that it is a false positive match, requires computing the probability of a random feature falling within the descriptor range of l. This is estimated by considering all features generated from the original tour, and finding the fraction which fall within the descriptor range of l, whilst in fact not belonging to the corresponding feature track.

The probability $p(S=1|N=0)$ that a random neighbour feature f_n will fall within the spatial arrangement range of a neighbour landmark l_n, can be computed by dividing the area occupied by this spatial range by the total possible area within which a neighbour feature can be found.

For an image of dimensions $w \times h$,

$$p(S=1|N=0) = \frac{(\theta_{\max} - \theta_{\min})(\delta_{\max} - \delta_{\min})}{2\pi \times w \times h} \quad (3)$$

Finally, computing $p(N|L=0)$ requires marginalizing over all locations in the topology. For a location i in which neighbour landmark l_n is present, $p(N=1|L=0) = 1/|\mathbf{f}_i|$, where $|\mathbf{f}_i|$ represents the total number of features present in the image representing location i. This is because only one feature out of the set $|\mathbf{f}_i|$ corresponds to the correct neighbour landmark, regardless of whether landmark l is correctly matched. For a location j in which neighbour landmark l_n is not present, $p(N=1|L=0)=0$, because a true positive neighbour match is not possible without the presence of l_n in the image. Therefore, for global localization with equal prior probabilities for each location,

$$p(N=1|L=0) = \sum_{x\in\mathbf{x}} p(x)p(N=1|x,L=0) = \frac{1}{|\mathbf{x}|}\sum_{x\in\mathbf{x'}}\frac{1}{|\mathbf{f}_x|} \quad (4)$$

Here, $|\mathbf{x}|$ represents to total number of nodes in the topology, and $\mathbf{x}'$ represents all locations whose images contain landmark l.

The final stage in matching features to landmarks is to choose which neighbour landmarks **n** for each landmark, should be considered for an attempted neighbour match. Requiring many neighbours to be matched to increases discriminative power, but significantly reduces computational efficiently, and the problem explodes to become an exhaustive graph matching algorithm. As such, the top k neighbours which co-occur most frequently with the landmark are chosen as candidate neighbour matches, and if one of these matches if verified then the belief of variable N in the Bayesian network is set to 1. The frequency of co-occurrence is computed by dividing the number of nodal images in which both the landmark and its neighbour are present, by the total number of nodal images in which the landmark is present.

V. Online Learning of Environment

One major benefit of extracting real-world landmarks and treating them independently is that the landmark properties can be updated over time as the robot explores its environment. Based upon the initial tour of the robot, for every nodal image that a landmark was extracted from, the probability that a landmark will occur if the robot is at this node, would be calculated as 1. However, in reality, due to different viewpoints and illumination conditions in subsequent tours, the landmarks are likely to be matched only a fraction of the time. Thus, a co-occurrence probability, $p(L=1|x=1)$ is computed between landmark l and node x, by dividing the number of times that node x is the matched node whilst landmark l is also matched, by the total number of times that node x is matched. This probability is updated every time node x is matched, and this is an important parameter in weighting landmark contributions in the image similarity measure, such that those landmarks which are expected to occur frequently are given more importance.

VI. Computing Location Similarities

With the match probabilities computed for each landmark, a similarity can now be computed between a query image and every nodal image in the topological map. For every feature in the query image, a match is considered to every landmark in memory. If the landmark descriptor, together with a neighbour descriptor and location, are all within the correct ranges, a match is recorded alongside a corresponding probability that this is a true positive match, as from Equation (1). For each location x in the topology, a similarity can be computed between the landmarks present in the query image, $l \in \mathbf{l}_q$, and those present in the map image, $l \in \mathbf{l}_x$. This similarity function $S(\mathbf{l}_q, \mathbf{l}_x)$ sums all the probabilities of landmark matches between the two images, with each weighted by the co-occurrence value $p(L=1|z_q=z_x)$, and divides this by the hypothetical value of the summation if all landmarks in $\mathbf{l}_x$ were matched with a true positive probability of 1. η is a normalizing factor, such that $p(\mathbf{l}_q|x)$ sums to 1 over all locations x.

$$p(z_q = z_x) \approx \eta S(q,x) = \eta \frac{\sum_{l\in\mathbf{l}_x} P(L=1|E_L=1)P(L=1|z_q=z_x)}{\sum_{l\in\mathbf{l}_x} P(L=1|z_q=z_x)} \quad (5)$$

VII. Optimization of Landmark Tracks

Currently, the landmarks generated in section III represent feature tracks that extend until a feature match is no longer found across adjacent images. However, this provides a suboptimal representation of the environment. Longer feature tracks will exhibit larger variance in their descriptor, together with larger variances in spatial relationships with neighbouring landmarks, and hence will be susceptible to greater false positive matches. However, the presence of longer feature tracks also reduces the processing time required, as fewer landmarks are formed to which feature matches are attempted. As such, it is possible to optimize the track lengths with respect to a required localization performance, whilst maintaining an acceptable frame rate.

A ratio threshold, r_x, is defined as the minimum ratio between the expected scene similarity of a true positive match between query image q and nodal image x, and the expected scene similarity of a false positive match between q and nodal image y, given that the robot location z_q does in fact correspond to location z_x in the map.

$$r_x = \frac{\mathrm{E}[S(\mathbf{I}_q, \mathbf{I}_x) \mid z_q = z_x]}{\max_{y \neq x} \mathrm{E}[S(\mathbf{I}_q, \mathbf{I}_y) \mid z_q = z_x]} \tag{6}$$

To compute the expected similarity for a true positive match, landmark matches are assumed to be conditionally independent given the matched image, and the contribution of each landmark to the similarity score is thus weighted by the probability that the landmark evidence will occur in the image, $p(L=1 \mid x=1)$.

$$\mathrm{E}[S(\mathbf{I}_q, \mathbf{I}_x) \mid z_q = z_x] = \eta \frac{\sum_{l \in \mathbf{I}_x} P(L=1 \mid E_L=1) P(L=1 \mid X=1)}{\sum_{l \in \mathbf{I}_x} P(L=1 \mid X=1)} \tag{7}$$

To compute the expected similarity for a false positive match, the contributions of each landmark are weighted by the probability of a false positive match to that landmark, $p(E_L=1 \mid L=0)$.

$$\mathrm{E}[S(\mathbf{I}_q, \mathbf{I}_y) \mid z_q = z_x] = \eta \frac{\sum_{l \in \mathbf{I}_x} P(L=1 \mid E_L=1) P(E_L=1 \mid L=0)}{\sum_{l \in \mathbf{I}_x} P(L=1 \mid X=1)} \tag{8}$$

The variable r is now defined as the overall minimum value of r_x across all nodal images x, and the user-defined variable r_{min} is defined as the minimum allowed value of r. In order to establish r to be greater than r_{min}, the following process is carried out. r is computed for the original map generated by the robot's tour. If $r < r_{min}$, each landmark in the nodal image x for which r_{min} is associated is ranked in order of its probability of a false positive match. The landmark with the highest value of this probability is split into two landmarks, each taking half the features in the feature track. r is then updated, and the process continues until $r > r_{min}$. In this way, as the landmarks are divided up, the likelihood of a false positive match between the query image and all nodal images is decreased to an acceptable level. Specifying the value of r is a compromise between localization accuracy and frame rate and can be adjusted to suit the particular needs of the system.

VIII. Experimental Results

Having built up the topological environment by manually driving a mobile robot along a path, experiments were then carried out to investigate the localization performance. The robot was driven along subsequent tours that deviated to the side of the original path by about 0.5 m, to test the robustness to slight viewpoint changes. Equation (5) was used to compute the most likely location within the map corresponding to the current image from the robot, and the overall percentage of correctly identified locations was recorded. Different path lengths of 20m, 50m and 100m were tested to investigate how well the proposed method scales with the environment.

Figure 4 shows the effect of online learning and updating of the co-occurrence probabilities of landmarks and nodal images. For each subsequent path that the robot was driven along, these probabilities were adjusted as in section V, allowing more weight to be given to frequently occurring landmarks in the similarity measure. This improves the localization performance because infrequently occurring landmarks which had a false positive match would not affect the similarity measure significantly, as they were not expected to be matched in the first place.

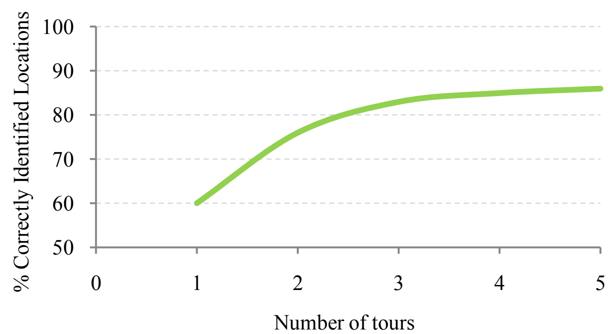

Fig. 4. Percentage of correctly identified rooms as the number of tours by the robot increases. Co-occurrence probabilities of landmarks and nodal images are updated with each subsequent tour, and landmark weights in the image similarity measure are adjusted accordingly. Each tour is 100m in length.

The localization performance was then compared to results generated using the traditional topological method [10]. In the comparative approach, features extracted from a query image are matched to all features from each individual image along the tour, by comparing the distances of the two closest feature descriptors, and hence discarding matches that are not confident. False positive feature matches are then pruned by estimating the epipolar geometry between the two images. In the traditional approach, inserting further nodes into the topology will significantly increase the processing time required to attempt a match to the new features. However, with the method proposed in this paper, an extra node is likely to contain features that belong to the already existing landmarks, and so the additional computational overhead is far less. The map density and value of r_{min} were adjusted, for each tour length, such that the frame rate of localization in this method was equal to the frame rate in the traditional topology, at around 3 frames per second. The results of this are shown in figure 5, where the map density is far greater in the proposed method, and additionally scales well with the environment.

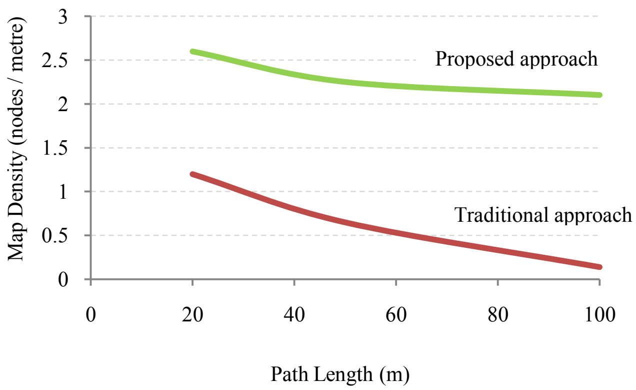

Fig. 5. Comparison of map density between the approach presented in this paper with the traditional topological approach, both operating at 3 frames per second.

The ability to create a greater map density, without decreasing frame rate, is the key advantage of this method over competing techniques. However, the overall localization accuracy is also improved in this method compared with [10], due to the appropriate weighting of informative landmarks and the ability the update the map through multiple tours. Figure 6 compares the localization rate of the two methods, again with both operating at around 3 frames per second.

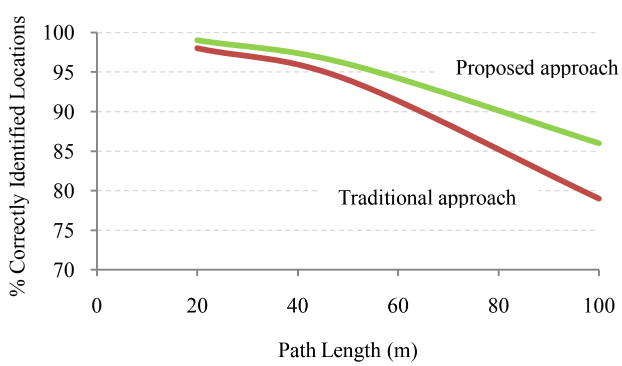

Fig. 6. Comparison of correct localization rate between the approach presented in this paper, and three alternative state-of-the-art approaches, at a frame rate of 3 frames per second.

IX. Conclusions

In this paper, a new approach to global topological localization has been presented. Traditional approaches match a query image to independent nodal images in a map, such that increasing the map density requires matching to a larger number of images, hence significantly reducing the frame rate. This method matches a query image to independent landmarks that have been tracked across multiple adjacent nodal images. As such, inserting further nodal images does not significantly reduce the frame rate, as the landmarks present in the additional images are likely to already be present in the map. By learning each landmark independently, statistics can be updated online which can assign greater importance to landmarks that are expected to occur more frequently in a nodal image, allowing for a greater location recognition rate than methods based upon a single training tour of the environment. Further proposed work may involve extending the online learning component to update landmark descriptor properties and the spatial properties of neighbouring landmarks. Additionally, the task of localizing in dynamic environments will benefit from this approach, as landmarks that no longer appear in subsequent tours by the robot can be filtered out, whilst new landmarks that appear can be gradually introduced into the map.

References

[1] S. Se, D. Lowe and J. Little, “Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks”, in *International Journal of Robotics Research*, 2002, 21(8), pp. 735-758

[2] A. Davison, I. Reid, N. Molton and O. Stasse, “MonoSLAM: Real-Time Single Camera SLAM”, in *Trans. PAMI*, 2007

[3] J. Kosecka and F. Li, “Vision Based Topological Markov Localization”, in *Proc. ICRA*, 2004

[4] M. M. Ullah, A. Pronobis, B. Caputo, J. Luo, P. Jensfelt and H. I. Christensen, “Towards Robust Place Recognition for Robot Localization”, in *Proc. ICRA*, 2008

[5] A. Georgiev, P. K. Allen, “Localization methods for a mobile robot in urban environments”, in *IEEE Trans. Robotics*, 2004, 20(5), pp. 851-864

[6] F. Dellaert, D. Fox, W. Burgard and S. Thrun, “Monte Carlo Localization for Mobile Robots”, in *Proc. ICRA*, 1999

[7] K. Grauman and T. Darrell, “Efficient image matching with distributions of local invariant features”, in *Proc. CVPR*, 2005

[8] J. Sivic and A. Zisserman, “Video Google: a text retrieval approach to object matching in videos”, in *Proc. ICCV*, 2003

[9] K. Ni and K. Kannan, “Epitomic Location Recognition”, in *Proc. CVPR*, 2008

[10] A. Ramisa, A. Tapua, R. L. de Mantaras and R. Toledo, “Mobile Robot Localization using Panoramic Vision and Combinations of Feature Region Detectors”, in *Proc. ICRA*, 2008

[11] C. Weiss, H. Tamimi, A. Masselli and A. Zell, “A Hybrid Approach for Vision-based Outdoor Robot Localization Using Global and Local Image Features”, in *Proc. IROS*, 2007

[12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. Van Gool, “A comparison of affine reion detectors” in *IJCV* 65(1/2), pp. 43-72, 2005

[13] D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, in *IJCV* 60(2), pp. 91-110, 2004

[14] A. Ascani, . Frontoni, A Mancini and P. Zingaretti, “Feature group matching for appearance-based localization”, in *Proc. IROS* 2008

[15] R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision”, Cambridge University Press, 2003

[16] S. Frintrop and A. B. Cremers, “Visual Landmark Generation and Redection with a Single Feature Per Frame”, in *Proc. ICRA* 2010

[17] L. Faying and J. Kosecka, “Probabilistic location recognition using reduced feature set”, in *Proc. ICRA* 2006

[18] E. Johns and G.-Z. Yang, “Scene Association for Mobile Robot Navigation”, in *Proc. IROS* 2010

[19] J. Kosecka and J. Xiaolong Yang, “Global localization and relative pose estimation based on scale-invariant features”, in *Proc. ICPR* 2004