Global Localization and Relative Positioning Based on Scale-Invariant Keypoints

Jana Košecká, Fayin Li and Xialong Yang

Computer Science Department,George Mason University, Fairfax, VA 22030

Abstract

The localization capability of a mobile robot is central to basic navigation and map building tasks. We describe a probabilistic environment model which facilitates global localization scheme by means of location recognition. In the exploration stage the environment is partitioned into a class of locations, each characterized by a set of scale-invariant keypoints. The descriptors associated with these keypoints can be robustly matched despite changes in contrast, scale and viewpoint. We demonstrate the efficacy of these features for location recognition, where given a new view the most likely location from which this view came is determined. The misclassifications due to dynamic changes in the environment or inherent appearance ambiguities are overcome by exploiting neighborhood relationships captured by a Hidden Markov Model. We report the recognition performance of this approach in an indoor environment consisting of eighteen locations and discuss the suitability of this approach for a more general class of recognition problems. Once the most likely location has been determined we demonstrate how to robustly compute the relative pose between the representative view and the current view, despite the partial absence of intrinsic parameters of the camera.

1 Introduction and Related Work

The two main instances of mobile robot localization problem are the continuous pose maintenance problem and the global localization also known as 'robot kidnapping' problem. While the successful solution to the localization problem requires addressing both, here we concentrate only on the global localization aspect. The problem of vision-based global localization shares many aspects of object recognition and hence is amenable to use of similar methodologies. While several instances of vision-based localization have been successfully solved in smaller scale

Email address: {kosecka,fli,xyang}@cs.gmu.edu (Jana Košecká, Fayin Li and Xialong Yang).

environments (1; 2; 3; 4), the applicability of these methods to large dynamically changing environment poses additional challenges and calls for alternative models. The methods for localization vary in the choice of features and the environment model. The two main components of the environment model are the descriptors chosen to represent an image and the representation of changes in image appearance as a function of viewpoint. Similarly as in the case of object recognition, both global and local image descriptors have been considered. The class of global image descriptors consider the entire image as a point in the high-dimensional space and model the changes in appearance as a function of viewpoint using subspace methods (5). Given the subspace representation the pose of the camera was typically obtained by spline interpolation method, exploiting the continuity of the mapping between the object appearance and continuously changing viewpoint. Robust versions of these methods have been applied in the robot localization using omnidirectional cameras (1). Alternative global representations proposed in the past include responses to banks of filters (6), multi-dimensional histograms (7; 8) or orientation histograms (9). These types of global image descriptors integrated the spatial image information and enabled classification of views into coarser classes (e.g. corridors, open areas), yielding only qualitative localization. In the case of local methods, the image is represented in terms of localized image regions, which can be reliably detected. The representatives of local image descriptors include affine or rotationally invariant features (10; 11) or local Fourier transforms of salient image regions (12). Due to the locality of these image features, the recognition is naturally prone to large amounts of clutter and occlusions. The sparser set of descriptors were in case of both global and local methods, typically obtained by principal component analysis or various clustering techniques.

Our approach is motivated by the recent advances in object recognition using local scale invariant features proposed by (10) and adopts the strategy for localization by means of location recognition. The image sequence acquired by a robot during the exploration is first partitioned to individual locations. The locations correspond to the regions of the space across which the features can be matched successfully. Each location is represented by a set of model views and their associated scale-invariant features. In the first localization stage, the current view is classified as belonging to one of the locations using standard voting approach. In the second stage we exploit the knowledge about neighborhood relationships between individual locations captured by Hidden Markov Model (HMM) and demonstrate an improvement in the overall recognition rate. The main contribution of this stage of the presented work is the instantiation of the Hidden Markov Model in the context of this problem and demonstration of an improvement in the overall recognition rate. This step is essential particularly in the case of large scale environments which often contain uninformative regions, violating the continuity the of the mapping between the environment appearance and camera pose. In such case imposing a discrete structure on the space of continuous observations enables us to overcome these difficulties while maintaining a high recognition rate. Once the most likely view has been determined we will show a simplified method for computing the rel-

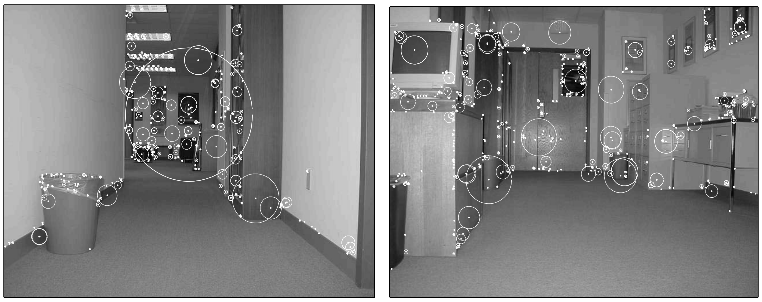


Fig. 1. The circle center represents the keypoint's location and the radius the keypoint's scale.

ative pose of the robot with respect to model view in the absence of the focal length of the camera. This second stage will then enable local metric localization given the model.

2 Scale-Invariant Features

The use of local features and their associated descriptors in the context of object recognition has been demonstrated successfully by several researchers in the past (13; 14; 15). In this paper we examine the effectiveness of scale-invariant (SIFT) features proposed by D. Lowe (10). The SIFT features correspond to highly distinguishable image locations which can be detected efficiently and have been shown to be stable across wide variations of viewpoint and scale. Such image locations are detected by searching for peaks in the image $D(x, y, \sigma)$ which is obtained by taking a difference of two neighboring images in the scale space

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma).$$

The image scale space $L(x, y, \sigma)$ is first build by convolving the image with Gaussian kernel with varying σ, such that at particular σ, $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$. Candidate feature locations are obtained by searching for local maxima and minima of $D(x, y, \sigma)$. In the second stage the detected peaks with low contrast or poor localization are discarded. More detailed discussion about enforcing the separation between the features, sampling of the scale space and improvement in feature localization can be found in (10; 16). Once the location and scale have been assigned to candidate keypoints, the dominant orientation is computed by determining the peaks in the orientation histogram of its local neighborhood weighted by the gradient magnitude. The keypoint descriptor is then formed by computing local orientation histograms (with 8 bin resolution) for each element of a 4×4 grid overlayed over 16×16 neighborhood of the point. This yields 128 dimensional feature vector which is normalized to unit length in order to reduce the sensitivity to image contrast and brightness changes in the matching stage. Figure 1 shows the keypoints found in the example images in our environment. In the reported experi-

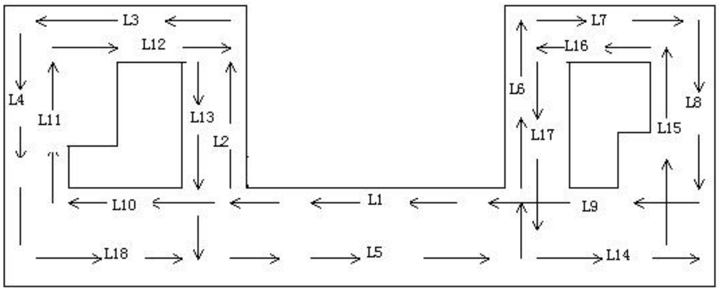


Fig. 2. The map on the fourth floor of our building. The arrows correspond to the heading of the robot and the labels represent individual locations.

ments the number of features detected in an image of size 480×640 varies between 10 to 1000. In many instances this relatively low number of keypoints, is due to the fact that in indoors environments many images have small number of textured regions. Note that the detected SIFT features correspond to distinguishable image regions and include both point features as well as regions along line segments.

3 Environment Model

The environment model, which we will use to test our localization method is obtained in the exploration stage. Given a temporally sub-sampled sequence acquired during the exploration (images were taken approximately every 2-3 meters), the sequence is partitioned into 18 different locations. The exploration route can be seen in Figure 3. Different locations in our model correspond to hallways, sections of corridors and meeting rooms approached at different headings. In the current experiment, the environment is mostly comprised of network of rectangular corridors and hallways which are typically traversed with four possible headings (N, S, W, E). The deviations from these headings can be handled as long as there is a sufficient overlap between the model views acquired during the exploration and current views. In case the current view cannot be matched successfully, a new location is added to the model. The number of views per location vary between 8 to 20 depending on the appearance variation within the location. The transitions between the locations occur either at places where navigation decisions have to be made or when the appearance of the location changes suddenly. The transitions between individual locations are determined depending on the number of features which can be successfully matched between the successive frames. These are depicted in Figure **??** for a sequence captured by a still digital camera along the path which visited all eighteen locations (some of them twice) and for a video sub-sequence along a path which visited three locations. The transitions between individual locations are marked by the peaks in the graph, corresponding to new locations. In order to obtain a more compact representation of each location a number of representative views is chosen per location, each characterized by a set of SIFT features. The sparsity of the model is directly related to the capability of matching SIFT

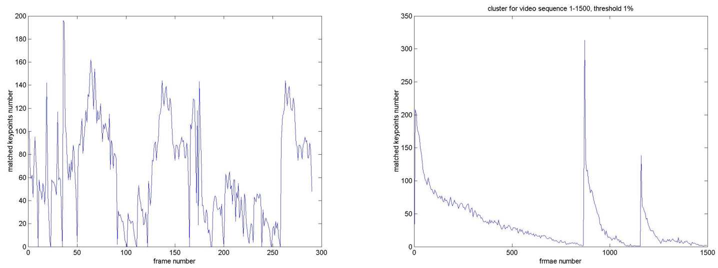


Fig. 3. The number of keypoints matched between consecutive views for the sequence comprised of 18 locations (snapshot was taken every 2-3 meters) captured by a digital camera (left); the number of keypoints matched between the first and i-th view for a video sequence comprised of 3 locations (right).

Fig. 4. Examples of representative views of 14 out of 18 locations.

features in the presence of larger variations in scale. The number of representative views varied between one to four per location and was obtained by regular sampling of the views belonging to individual locations. Examples of representative views associated with individual locations are depicted in Figure 4.

4 Location recognition

The environment model obtained in the previous section consists of a database of model views [1]. The i-th location in the model, with $i = 1, \ldots N$ is represented by n views $I_1^i, \ldots, I_n^i$ with $n \in \{1, 2, 3, 4\}$ and each view is represented by a set of SIFT features $\{S_k(I_j^i)\}$, where k is the number of features. In the initial stage we tested the location recognition by using a simple voting scheme. Given a new query image Q and its associated keypoints $\{S_l(Q)\}$ a set of corresponding keypoints between

[1] It is our intention to attain a representation of location in terms of views (as opposed to some abstract features) in order to facilitate relative positioning tasks in the later metric localization stage.

(# of views)	#1 (250)	#2 (134)	#3 (130)
one	84%	46%	44%
two	97.6%	68%	66%
four	100%	82%	83%

Table 1
Recognition performance for one training and two test sequences in terms of % of correctly classified views as a function of number of representative views.

Q and each model view I_j^i, $\{\mathcal{C}(Q, I_j^i)\}$, is first computed. The correspondence is determined by matching each keypoint in $\{S_l(Q)\}$ against the database of $\{S_k(I_j^i)\}$ keypoints and choosing the nearest neighbor based on the Euclidean distance between two descriptors. We only consider point matches with high discrimination capability, whose nearest neighbor is at least 0.6 times closer then the second nearest neighbor. More detailed justification behind the choice of this threshold can be found in (10). In the subsequent voting scheme we determine the location whose keypoints were most frequently classified as nearest neighbors. The location where the query image Q came from is then determined based on the number of successfully matched points among all model views

$$C(i) = \max_j |\{\mathcal{C}(Q, I_j^i)\}| \text{ and } [l, num] = \max_i C(i)$$

where l is the index of location with maximum number num of matched keypoints. Table 1 shows the location recognition results as a function of number of representative views per location on the training sequence of 250 views and two test sequences of 134 and 130 images each. All three sequences were sparse with images taken 2-3 meters apart. The two test sequences were taken at different days and times of day, exhibiting larger deviations from the path traversed during the training. Despite a large number of representative views per location relatively poor performance on the second and third test sequence was due to several changes in the environment between the training and testing stage. In 5 out of 18 locations several objects were moved or misplaced. Examples of dynamic changes can be seen in Figure 5. The poorer performance due to dynamic changes is not surprising, since the most discriminative SIFT features often belong to objects some of which are not inherent to particular locations. In the next section we describe how to resolve these issues by modelling the neighborhood relationships between individual locations.

5 Modelling spatial relationships between locations

We propose to resolve these difficulties by incorporating additional knowledge about neighborhood relationships between individual locations. The rationale behind this choice is that despite the presence of ambiguities in recognition of indi-

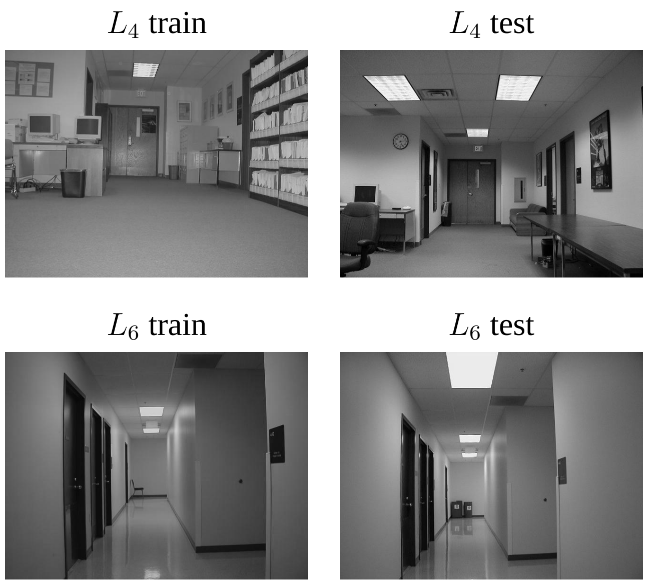
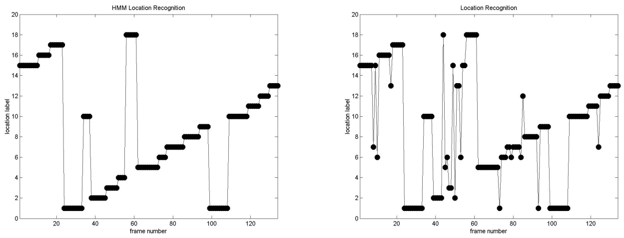


Fig. 5. Changes in the appearance of location L_4 and L_6 between the training and testing. In the left image pair the bookshelve was replaced by a table and couch and in the right pair recycling bins were removed.

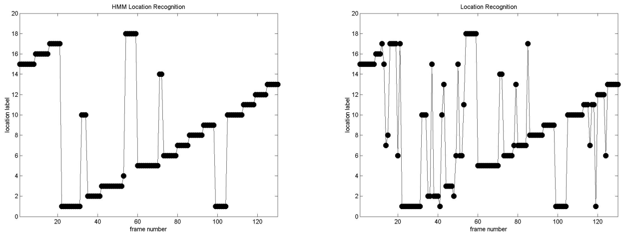
vidual views the temporal context should be instrumental in resolving them. The use of temporal context is motivated by the work of (17) which addresses the place recognition problem in the context of wearable computing application. The temporal context is determined by spatial relationships between individual locations and is modelled by a Hidden Markov Model (HMM). In this model the states correspond to individual locations and the transition function determines the probability of transition from one state to another. Since the locations cannot be observed directly each location is characterized by The most likely location is at each instance of time obtained by maximizing the conditional probability $P(L_t = l_i | o_{1:t})$ of being at time t and location l_i given the available observations up to time t. The location likelihood can be estimated recursively using the following formula

$$P(L_t = l_i | o_{1 \ldots t}) \propto p(o_t | L_t = l_i) P(L_t = l_i | o_{1:t-1}) \tag{1}$$

where $p(o_t | L_t = l_i)$ is the observation likelihood, characterizing how likely is the observation o_t at time t to come from location l_i. The choice of observation likelihood depends on the available observations and the matching criterion. When local descriptors are used as observations, several such choices have been proposed in the context of probabilistic approaches to object recognition (18; 19). The proposed likelihood functions properly accounted for the density and spatial arrangements of features and improved overall recognition rate. In the context of global image descriptors the locations were modelled in terms of Gaussian mixtures proposed in (17). Since the location recognition problem is notably simpler then the object recognition problem due to occlusions and clutter not being some prominent, we used a simpler form of the likelihood function. The conditional probability $p(o_t | L_t = l_i)$ that a query image Q_t at time t characterized by an observation



Seq. 2 with and without HMM



Seq. 3 with and without HMM

Fig. 6. Classification results for Sequence 2 and Sequence 3 with and without considering the spatial relationships. The black circles correspond to the labels of most likely locations.

$o_t = \{S_l(Q_t)\}$ came from certain location, is directly related to the cardinality of the correspondence set $C(i)$ and the distance between individual descriptors. Lets denote the set of descriptors associated with the query view $\{g_k^Q\}$ and the $\{g_k^i\}$ set of descriptors of the ith model view. We then denote

$$p(o_t|L_t = l_i) = p(\{g_k^Q\}|L_t = l_i) = 1 - (\Pi_{k=1}^{n}(1 - exp(-\frac{\alpha_k^i}{2\sigma^2}))$$

where α_k^i is so called strangeness parameter and is defined as

$$\alpha_k^i = \frac{min_{g_j \in S_i}(\|g_k^Q - g_j\|)}{min_{g_j \notin S_i}(\|g_k^Q - g_j\|)}. \qquad (2)$$

It is the ratio of minimal intra-distance within the class and minimal inter-distance to location putative label l. If α_k^i is greater than 1, the feature g_k^Q is not contributing to classification of Q as label i. In order to explicitly incorporate the location neighborhood relationships, the second term of equation (1) can be further decomposed

$$P(L_t = l_i|o_{1:t-1}) = \sum_{j}^{N} A(i,j)P(L_{t-1} = l_j|o_{1:t-1}) \qquad (3)$$

where N is the total number of locations and $A(i,j) = P(L_t = l_i|L_t = l_j)$ is the probability of two locations being adjacent. In the presence of a transition between two locations the corresponding entry of A was assigned a unit value and in the final stage all the rows of the matrix were normalized. The results of location recognition employing this model are in Figure 6. The recognition rate for Sequence 2 was 96.3% and for Sequence 3 it was 95.4%. The location label assigned to each image is the one with the highest probability. While in both cases some images were misclassified the overall recognition rates are an improvement compared to the rates reported in Table 1. Despite the classification errors in Sequence 2, the order of visited locations was correctly determined. For Sequence 3, where we exhibited some intentional deviations between the path taken during training and testing, the classification of location 14 was incorrect. The effect of HMM model can be examined by making all the probabilities in the transition matrix A uniform essentially neglecting the knowledge of location neighborhood relationships. The assigned location labels for this case are in the right column of Figure 6, with noticeably degraded recognition performance.

6 Pose Estimation

Once the most likely location and best matched view has been found we can compute the relative displacement between the current view and model view.

The current view and the matched model view are related by a rigid body displacement $g = (R, T)$ represented by a rotation $R \in SO(3)$ and translation $T = [t_x, t_y, t_z]^T \in \mathbb{R}^3$. Provided that the camera is calibrated, g can be estimated from the epipolar geometry between the two views. This recovery problem can be further simplified taking into account the fact that the motion of the robot is restricted to a plane. Here we outline an algorithm for this special case and demonstrate how to recover the displacement in case of unknown focal length. The case of general motion and unknown focal length was studied by (20) and the solution for the case of planar motion case has been proposed by (21) in the context of uncalibrated stereo. Here we demonstrate a slightly different, more concise solution to the problem. Consider the perspective camera projection model, where 3D coordinates of point $\mathbf{X} = [X, Y, Z]^T$ are related to their image projections $\mathbf{x} = [x, y, 1]^T$ by an unknown scale λ; $\lambda \mathbf{x} = \mathbf{X}$. In case the camera is calibrated the two views of the scene are related by $\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T$, where $(R, T) \in SE(3)$ is a rigid body transformation and λ_1 and λ_2 are the unknown depths with respect to individual camera frames. After elimination of the unknown scales from the above equation, the relationship between the two views is captured by so-called epipolar constraint

$$\mathbf{x}_2^T \widehat{T} R \mathbf{x}_1 = \mathbf{x}_2^T E \mathbf{x}_1 = 0, \tag{4}$$

where $E = \widehat{T}R$ is the essential matrix [2] In case of planar motion, assuming translation in $x - z$ plane and rotation around $y-$axis by an angle θ, the essential matrix

[2] $\widehat{T}$ denotes a 3×3 skew symmetric matrix associated with vector T.

has the following sparse form

$$E = \begin{bmatrix} 0 & -t_z & 0 \\ t_z c\theta + t_1 s\theta & 0 & t_z s\theta - t_1 c\theta \\ 0 & t_x & 0 \end{bmatrix} \tag{5}$$

where $s\theta (c\theta)$ denote $\sin\theta (\cos\theta)$ respectively. Given at least four point correspondences, the elements of the essential matrix $[e_1, e_2, e_3, e_4]^T$ can be obtained as a least squares solution of a system of homogeneous equations of the form (4). Once the essential matrix E has been recovered, the four different solutions for θ and $T = \pm[t_x, 0, t_z]$ can be obtained (using basic trigonometry) directly from the parametrization of the essential matrix (5). The physically correct solution is then obtained using the positive depth constraint. In the case of unknown focal length the two views are related by so called fundamental matrix F

$$\tilde{\mathbf{x}}_2^T F \tilde{\mathbf{x}}_1 = 0 \text{ with } \mathbf{x} = K^{-1}\tilde{\mathbf{x}}. \tag{6}$$

The fundamental matrix F is in this special planar, partially calibrated case related to the essential matrix E as follows

$$F = K^{-T} E K^{-1} \text{ with } K = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

where f is the unknown focal length. The remaining intrinsic parameters are assumed to be known. In the planar motion case the matrix $F = [0, f_1, 0; f_2, 0, f_3; 0, f_4, 0]$ can be recovered from the homogeneous constraints of the form (6) given a minimum of four matched points. The extraction of the unknown motion parameters and the focal length f however is not straightforward, since the translation and the focal length appear in the parametrization of the matrix F in a multiplicative way. We propose to use additional constraints provided by so-called Kruppa's equations (22). It can be easily verified that a fundamental matrix F between the two views and the unknown intrinsic parameter matrix K satisfy the following constraint

$$FKK^T F^T = \lambda^2 \hat{e} KK^T \hat{e}^T \tag{8}$$

where $e = \frac{KT}{\|KT\|}$ is the epipole and λ is the unknown scale of the fundamental matrix. In our previous work (22) we have shown that for the special case of planar motion the above equation is satisfied if and only if $\lambda = 1$. Since F and $e = [-f_1, 0, f_4]^T$ can be estimated, the renormalized equation (8) yields following useful constraint on intrinsic parameters K

$$FKK^T F^T = \hat{e} KK^T \hat{e}^T. \tag{9}$$

Given the planar motion case, the middle entries of matrices on the left and right side of equation (9) yield a constraint on the focal length and the entries of the fundamental matrix

$$f_2^2 f^2 + f_3^2 = f_4^2 f^2 + f_1^2.$$

The solution for the focal length can then be directly obtained from the above equation as

$$f = \sqrt{\frac{f_1^2 - f_3^2}{f_2^2 - f_4^2}}. \tag{10}$$

Once f is computed, the relative displacement between the views can be obtained by the method outlined for the calibrated case. Additional care has to be taken in assuring that the detected matches do not come from a degenerate configuration. We have used RANSAC algorithm for the robust estimation of the pose between two views, with slightly modified sampling strategy. Figure 7 shows two examples of relative positioning with respect to two different representative views. The initial estimate of the motion and focal length is further refined by nonlinear minimization, where the total reprojection error of all the matched points in minimized

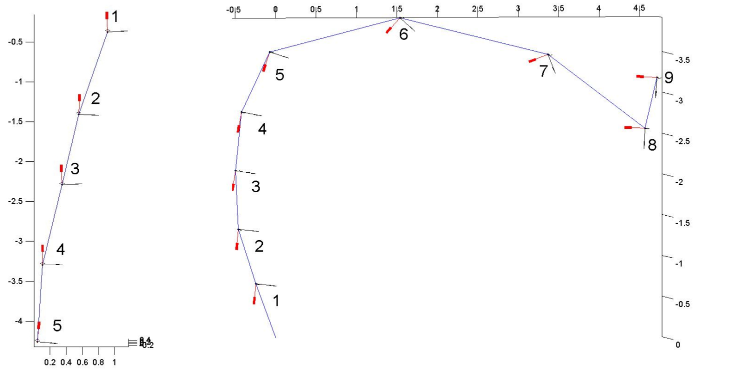
$$E(R,T,f) = min \sum_{i=1}^{n} \|\mathbf{x}^i - \pi([KR, KT]\mathbf{X}^i)\|^2 + \|\mathbf{x}_r^i - \pi(\mathbf{X}^i)\|^2, \tag{11}$$

where K is partially known matrix of intrinsic parameters, $\mathbf{x}^i$ and $\mathbf{x}_r^i$ are the matched SIFT features between the current view and the most likely reference view and $\mathbf{X}_i$ are 3D coordinates of points expressed with respect to the reference view. Note that 3D structure of the scene is estimated as well. The focal length estimates obtained for these examples are $f = 624.33$ and $f = 545.30$. The relative camera pose for individual views is represented in the figure by a coordinate frame.

7 Conclusions and Future Works

We have demonstrated the suitability and the discrimination capability of the scale-invariant SIFT features in the context of location recognition and global localization task. Although the matching and location recognition methods can be accomplished using an efficient and simple voting scheme, the recognition rate is affected by dynamic changes in the environment and inherent ambiguities in the appearance of individual locations. We have shown that these difficulties can be partially resolved by exploiting the neighborhood relationships between the locations captured by Hidden Markov Models.

Since the notion of location is not defined precisely and is merely inferred in the learning stage the presented method enables only qualitative global localization in terms of individual locations. Following the global localization we compute the relative pose of the robot with respect to the closest reference view (24) found



Location 1 Location 2

Location 1

Location 2

Fig. 7. Relative positioning experiments with respect to the representative views. Bottom: Query views along the path between the first view and the representative view for two different locations. Top: Recovered motions for two locations.

in the matching stage. This enables us to achieve metric localization with respect to the reference view, which can be followed by relative positioning tasks. More extensive experiments as well as integration with the exploration and navigation strategies on-board of mobile robot platform are currently underway.

8 Acknowledgements

The authors would like to thank D. Lowe for making available the code for detection of SIFT features.

References

[1] Artac, M., Jogan, M., Leonardis, A.: Mobile robot localization using an incremental eigenspace model. In: IEEE Conference of Robotics and Automation.

(2002) 1025 – 1030
[2] Gaspar, J., Winters, N., Santos-Victor, J.: Vision-based navigation and environmental representations with an omnidirectional camera. IEEE Transactions on Robotics and Automation (2000) 777–789
[3] Davidson, A., Murrray, D.: Simultaneous localization and map building using active vision. IEEE Transactions on PAMI **24** (2002) 865–880
[4] Se, S., Lowe, D., Little, J.: Global localization using distinctive visual features. In: Proc. of International Conference on Robots and Systems. (2002) 153–158
[5] Nayar, S., Nene, S., Murase, H.: Subspace methods for robot vision. IEEE Transactions on Robotics and Automation (**6**) 750–758
[6] Torralba, A., Sinha, P.: Recognizing indoor scenes. MIT AI Memo (2001)
[7] Schiele, B., Crowley, J.L.: Object recognition using multidimensional receptive field histograms. International Journal of Computer Vision (2000)
[8] H. Aoki, B.S., Pentland, A.: Recognizing places using image sequences. In: Conference on Perceptual User Interfaces, San Francisco (1998)
[9] Košecká, J., Zhou, L., Barber, P., Duric, Z.: Qualitative image based localization in indoors environments. In: IEEE Proceedings of CVPR. (2003) 3–8
[10] Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2004) to appear
[11] Wolf, J., Burgard, W., Burkhardt, H.: Using and image retrieval system for vision-based mobile robot localization. In: Proc. of the International Conference on Image and Video Retrieval (CIVR). (2003)
[12] Sims, R., Dudek, G.: Learning envirmental features for pose estimation. Image and Vision Computing **19** (2001) 733–739
[13] Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Transactions on PAMI **19** (1997) 530–534
[14] Selinger, A., Nelson, R.: A perceptual grouping hierarchy for appearance-based 3d object recognition. Computer Vision and Image Understanding (**76**) 83–92
[15] Lowe, D.: Object recognition from local scale invariant features. In: International Conference on Computer Vision. (1999) 1150–1157
[16] Brown, M., Lowe, D.: Invariant features from interest point groups. In: In Proceedings of BMVC, Cardiff, Wales. (2002) 656–665
[17] Torralba, A., Murphy, K., Freeman, W., Rubin, M.: Context-based vision system for place and object recognition. In: International Conference on Computer Vision. (2003)
[18] Pope, A., Lowe, D.: Probabilistic models of appearance for 3-d object recognition. International Journal of Computer Vision **40** (2000) 149 – 167
[19] Schmid, C.: A structured probabilistic model for recognition. In: Proceedings of CVPR, Kauai, Hawai. (1999) 485–490
[20] Sturm, P.: On focal length calibration from two views. In: Conference on CVPR, Kauai. Volume 2. (2001) 145–150
[21] Brooks, M., Agapito, L.D., Huyng, D., Baumela, L.: Towards robust metric reconstruction via a dynamic uncalibrated stereo. Image and Vision Comput-

ing **16** (1998) 989–1002
[22] Ma, Y., Vidal, R., Košecká, J., Sastry, S.: Kruppa's equation revisted: Degeneracy and renormalization. In: ECCV. (2000)
[23] Košecká, J., Yang, X.: (Global localization and relative positioning based on scale-invariant keypoints)
[24] Xang, X., Košecká, J.: Experiments in location recognition using scale-invariant sift features. Technical Report GMU-TR-2004-2, George Mason University (2004)