Visual localization using an optimal sampling of bags-of-features with entropy

Loic Lachèze and Ryad Benosman

Abstract— **This paper investigates the visual Localization of a mobile platform using a new sampling adaptive bag-of-features patches techniques. The method is specifically developed for the navigation of robots. It is based on the idea of an adaptive dense sampling of images using an optimal multilayer Quad-tree decomposition of the image driven by the quantity and homogeneity of the information contained within subpatches. Extracted patches will be of different sizes according to the covered zones in the image. Experimental results carried out on real images in the case of a navigation of a mobile robot using an omnidirectional camera are presented. The method of generating maps will be briefly introduced. Large amount of measures in different cases of noise and occlusion will be presented showing the robustness of the method.**

I. INTRODUCTION

Many authors have proposed various approaches to vision-based Localization systems in robotics. They mainly differ by the features they use to match images. Localization is a difficult task as in general it is asked to a mobile platform to start from an unknown location in an unknown environment and then simultaneously build a map of the environment and localise itself in the map. This issue is known as the Simultaneous Localization And Mapping (SLAM) problem. Navigation systems may use either topological or metric maps [15]. In topological maps, the relations between locations are learned and recognized [16], whereas in metric maps the introduction of a metric involves a higher precision of environment features and robot location [17], [18]. Methods merging both can be found in [19]. To ease the task, the environment is sometimes assumed known [25]. Some approaches use lines and edges to build a geometric model from images [22], in other cases the same features can be merged with ultrasound sensors [24] to be more robust. Several aspects of the environment can also be extracted from images like walls, doors and vacant spaces [21]. Other work rely on neural networks [23] to learn the location of the robot. There also exists methods applying a principal mathematical component analysis to learn landmarks [20]. Trajectories [26], [27] can also be used for Localization. Colorimetric methods can be in some cases valuable tools [28] specially in landmarks identification. Scale-invariant features are a promising technique, that can be used in many applications and can provide interesting results in estimating locations [37]. Depth information using stereo images can also be applied [30] sometimes also to retrieve the topology of scenes. There are approaches that rely on image-retrieval techniques to identify the position of robots [31], [32]. Finally Bayesian techniques can be found [33], [34], [35] in a wide variety of robot applications. Recently, bag-of-features representations have become popular, they are geometry free features based purely on characterising the statistics of local patch appearances. They are simple to use, robust, and allow good performance. The basic idea behind the method is to extract a set of local image patches from images. The patches are sampled and assigned a metric description. The resulting quantified descriptors give an implicit distribution description space that can be quantified using different methods. Most of the existing work differ mainly according to the way patches are sampled and then described. Patches can be selected using keypoints [1], [5], [4], [6], [7], [8], [9], [10], [11], dense sampling [3], [13], [14], SIFT based approaches [1], [6], [8], [10], filter based [3], [13] and raw patch based [4], [5], [7], [11]. Codebooks are then produced using k-means [1], [3], [11], [13] and agglomerative [4], [7] clustering. Sampling patches is surely the critical component of any bag-of-features method. Presented techniques are usually applied to recognize objects and are not really adapted for a task of robot navigation. They consider partial information from scenes, which in case of mobile robotics limits the robustness of visual loops. The presented method is driven by the idea that a robust localization needs all the information contained in images. This paper introduces a new technique of sampling images developped specifically for the navigation of robots. It is based on the idea of a dense sampling using an optimal multilayer decomposition of the image driven by the quantiy and homogeneity of the information contained within sub-patches. Extracted patches will be of different sizes according to the covered zone in the image. The paper is organised as follows, section one, will introduce the basic concepts used to treat images as a loose collection of independent patches explaining how patches are described, and how the most representative ones are selected. We will then present the sampling method to extract patches from images, based on a measure of entropy and image decomposition. We will then present experimental results in the case of a navigation of a robot using an omnidirectional camera. The method of generating maps will be briefly introduced, and finally a large amount of measures in different cases of noise and occlusion will be presented.

L. Lachèze is at ISIR , University of Pierre and Marie Curie, 75015 Paris, France loic.lacheze@lip6.fr

R. Benosman is with ISIR, University of Pierre and Marie Curie, 75015 Paris, France benosman@ccr.jussieu.fr

II. From images to patches

A. Characterizing texture

Texture can be measured using different approaches. In what follows we choose to use a measure similar to [36]. It relies on the computation of a histogram of the difference between the value of pixels of images. Given an image I, each value of its histogram of differences h_I is given by :

$$h_I(i) = \sum_{x,y,x',y' \in I}^{x \neq x' \vee y \neq y'} diff(I,x,y,x',y',i),\ i \in [0,255]$$

with

$$diff(I,x,y,x',y',i) = \begin{cases} 1 & \text{if } |I(x,y) - I(x',y')| = i \\ 0 & \text{else} \end{cases}$$

In a second stage, the histogram h_I is normalized, to ensure an invariance according to the size of I.

B. Generating codebooks

Let $F_z(I)$ be a function allowing the decomposition of an image I into several textured patches :

$$F_z(I) = z_0, z_1, ..., z_n \text{ with } I = \bigcup_{i=0}^{n} z_i$$

At this stage for a better understanding will consider that all z_i are squares with equal areas. Let $T = h_{z_0}, h_{z_1}, ...h_{z_n}$ be the set containing all texture descriptors of patches z_i of I. The idea is to sample T to reduce the number of descriptors to $m \leq n$. We then add to T a metric function expressed by $dist(h_{z_i}, h_{z_j})$ and a reference texture patch h_{ref}. The reference patch is set to a patch containing a single color, corresponding to a uniform area. In a second stage all the representation of patches contained in T are compared to h_{ref} and sorted, from the less to the more textured. The set T_s corresponding to the ordered set T becomes :

$$T_s = h_{ref}, h'_{z_0}, h'_{z_1}, ...h'_{z_n}$$
$$\text{with } dist(h_{ref}, h'_{z_i}) \leq dist(h_{ref}, h'_{z_j}) \text{ if } i < j$$

An example of T_s is given by figure.1, where for a better understanding a simple object is considered. The image is cut into equal square areas, the sorting appears clearly as patches are sorted according to their texture. The mahalanobis distance is used as a metric function and is set so for the rest of the paper.

At this point, T_s still contains too much information, it is then sampled into m equal areas. For each area, only the median patch is selected. The resulting selection gives the codebook V :

$$V = h_{ref}, h'_{z_0}, h'_{z_1}, ...h'_{z_m},\ V \subset T_s$$

that corresponds to the most representative patches. It is also possible to consider a much more complex sampling for example to extract more information from dense areas and less from empty ones, this should be adjusted according to the complexity of scenes. One other andvantage of this approach is that m can be set by the user and adjusted according to available robot memory.

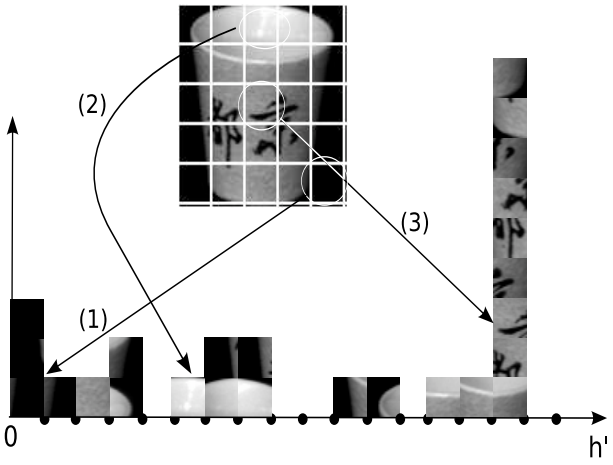

Fig. 1. Example of decomposing an image into patches. Extracted patches are compared to a uniform patch and sorted according to the complexity of their texture. The reference patch is set to 0 on axis h'. The mahalanobis distance is used as the metric of comparison. Example of patches from the less (1) textured to the more textured (3) are shown. The patch shown in (2) contains a light reflection that is assimilated to texture after normalization.

C. Decomposing images into known patches of V

Let I_{acq} be an acquired image, I_{acq} is decomposed into z_{acq_i} patches. Each computed patch must be compared to the content of V, we then set a function $Reco$ that transforms the patches of acquired images into patches of the codebook V :

$$Reco(z_{acq_i}, V) : if \exists j, h'_{z_j} \in V / dist(h'_{z_j}, h_{z_{acq_i}}) = 0$$

$$\text{then } h_{z_{acq_i}} = h'_{z_j} \text{ else } V = V \cup h_{z_{acq_i}}$$

In case a new patch is detected it is added to the codebook as a new entry. The acquired image I_{acq} is then codified using the patches of the codebook, the resulting image I_{cod_i} given by :

$$I_{cod_i} = Reco(z_{acq_0}, V), Reco(z_{acq_1}, V), ..., Reco(z_{acq_m}, V)$$

The acquired image of figure.2 shows an outdoor scene, the image is decomposed and expressed according to the codebook patches. Figure.3 presents different codifications

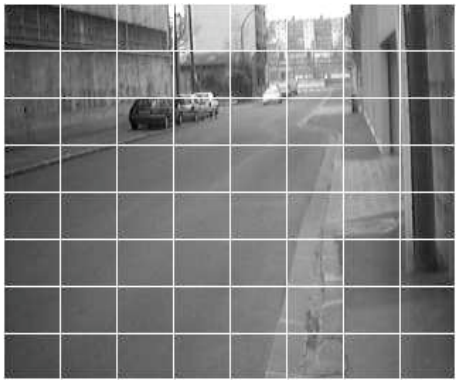

Fig. 2. Outdoor test image presented with patches of size 5×5 pixels.

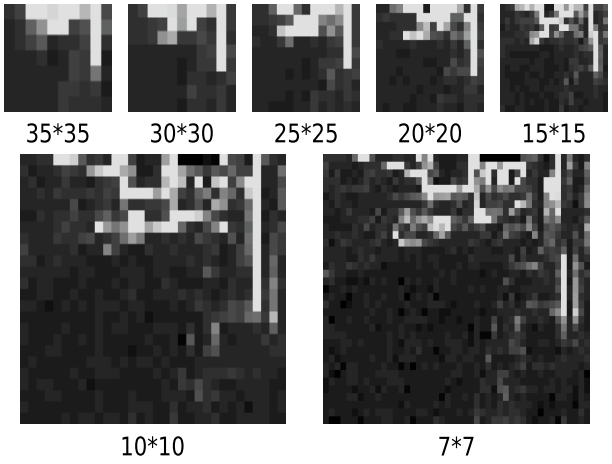

Fig. 3. Example of image codification of the image of figure.2 obtained for several values of patches varying from 35×35 to 7×7 pixels. The codebook has a maximal size of 32 patches.

according to the size of the considered patches. It is interesting to notice that there is a straight link between the size of patches and the accuracy of the description in this case.

D. Comparing codified images

A codified image I_{cod_i} is directly expressed using representative patch index as shown in figure.3. Comparing two codified images I_{cod_i} and I_{cod_j} is given by :

$$Comp(I_{cod_i}, I_{cod_j}) = \sum_{i=1}^{m} (I_{cod_i} - I_{cod_j})^2$$

Substracting images introduce a spatial dimension to the comparison, as due to the optimal patch decomposition two close images will theoretically have the same decomposition. Histogram of patches can also be used, but in the present case due to the spatial properties of the decomposition they would provide less interesting results.

E. Optimal decomposition of images.

The dense cutting images according to squares of equal areas introduce a major defect as it is not robust to translation. The decomposition being not unique, known images may then not be recognised. An efficient decomposition must produce an optimal and possibly a unique partitioning of images. In addition it would be interesting to produce less patches, but of variable size so that they can cover homogeneous texture zones

In order to achieve an optimal generation of patches, a quadtree-like algorithm is set up. Quadtree algorithm cut recursively images into subimages and so on. Starting for the initial image, each subimage is cut into four equal subimages. The idea is to use the same principle, but at the contrary of the regular quad-tree approach, the division of subimage will be driven by a measure of an entropy measure. The idea

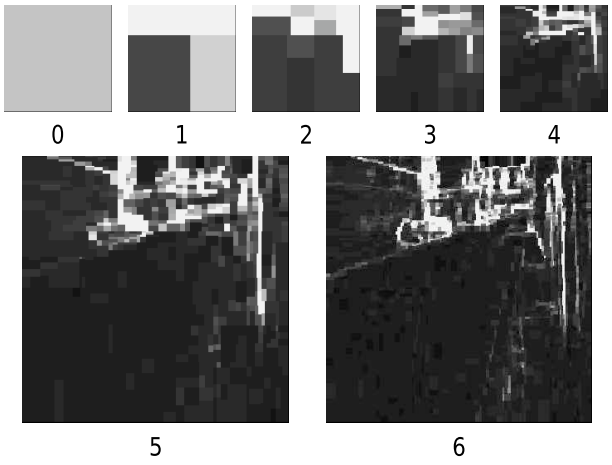

Fig. 4. Exemple of codification of a scene using the optimal method for every layer of figure.2. The codebook has a maximal size of 32 patches.

is to cut a subimage at the location were the difference of the quantity of information between possible subimages is minimal.

This quantity of information is given for an image I by :

$$H(I) = -\sum_{c=0}^{c=255} Occ(I = c) \log P(I = c)$$

with $Occ(I = c)$ the number of times the pixel value c appears in I, $P(c)$ is the probability of appearance of the grey value c within I.

To estimate the optimal point minimizing the variance of the distance between the information contained in the four subimages of I, the principle of integral images introduced in [38] is used.

Let $q(i,j)$ be the quantity of information of a pixel $I(i,j)$ with $q(i,j) = \log(P(I(i,j)))$.
We set the integral information of $I(x,y)$ as :

$$QI(x,y) = \sum_{i \leq x, j \leq y} q(i,j)$$

This sum is computed in one iteration on the whole image or subimage considered. We set $R(x,y-1)$ the integral quantity of information on the row x of height $y-1$. The principle of computation is presented in figure.5(A).

$$R(x,y-1) = QI(x,y-1) - Q(x-1,y-1)$$

Finally the integral quantity of information for a (x,y) (see figure.5(B)(C) is given by:

$$QI(x,y) = QI(x-1,y) + R(x,y-1) + q(i,j)$$

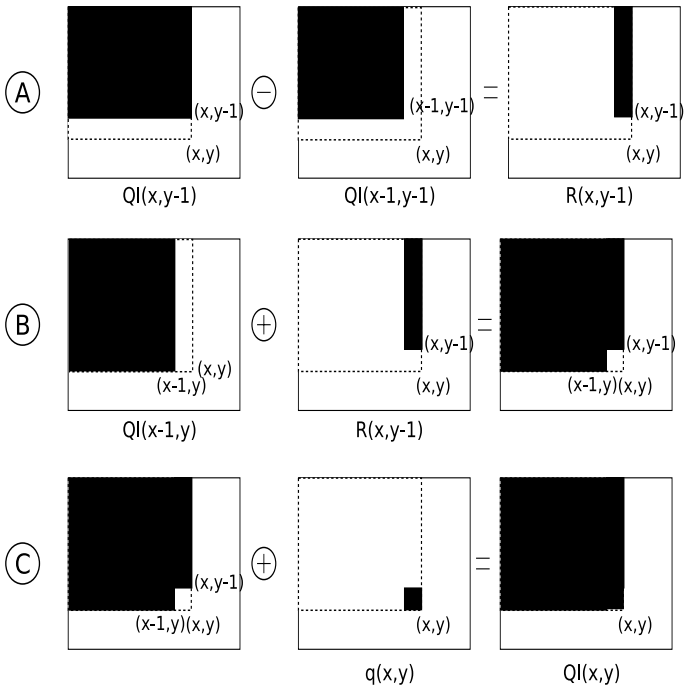

Fig. 5. Computation of $QI(x,y)$ using $QI(x-1,y-1)$, $QI(x,y-1)$, $QI(x-1,y)$ and $q(x,y)$. In (A) The computation of $R(x,y-1)$ and in (B) and (C) the final computation of $QI(x,y)$.

Once QI is computed the variance value within each pixel becomes implicit.

It is important to compute the mean value of the quantity of information contained within the four subimages. In the case where I is of size $m \times n$ we have for a cutting position $(x = m/2, y = n/2)$:

$$QI_m = QI(m,n)/4$$

As shown in figure.6, the quantity of information of each zone is :

$$QI_{11}(x,y) = QI(x,y)$$
$$QI_{12}(x,y) = QI(m,y) - QI(x,y)$$
$$QI_{21}(x,y) = QI(x,n) - QI(x,y)$$
$$QI_{22}(x,y) = QI(m,n) - QI_{21} - QI_{12} + QI_{11}$$

.

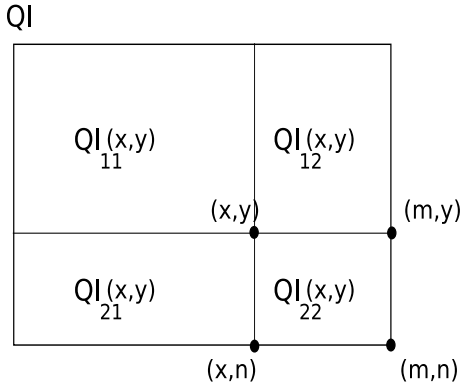

Fig. 6. Computation of the quantity of information contained in the four subimages of QI for a cutting position located at (x,y).

Finally the optimal (x,y) position is the one minimizing the following sum of differences:

$$\exists (x,y)/min_{x,y}(\sum_{a=1,b=1}^{a=2,b=2} (QI_m - QI_{ab}(x,y))^2)$$

An illustration of the algorithm is given in figure.4. It appears clearly that the image is decomposed more coherently. Comparing results of figure.3 of 10×10 squares with figure.4(5), visually shows a more precise decomposition for the optimal method due to the use of specific size patches. It is interesting to notice that features like edges and texture appear clearly. To illustrate the stability of the decomposition figure.7 presents the optimal patches for a simple object translated between the two images. One may check the stability of the method by noticing that in both cases patches cover the same zones despite the important translation of the considered object.

Fig. 7. Optimal generation of patches for a translated object. The object is covered with the same patches, covering the same areas and providing an equal decomposition of the image.

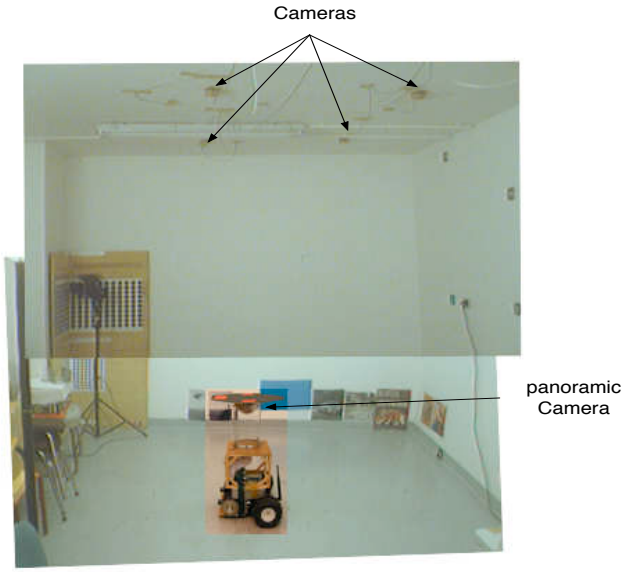

Fig. 8. Experimental environment, to ensure a ground truth, the position of the robot is estimated in real time by a camera network mounted on the ceiling of the room.

III. EXPERIMENTS

A. Experimental Settings

Experimental results are carried out using an autonomous differential drive robot, equipped with an omnidirectional catadioptric sensor mounted on top. The position of the robot is constantly estimated by a camera network mounted on top. This provides a ground truth measure of the estimated positions given by the navigation system of the robot (see figure.8). The robot explores randomly its environment, and creates a dense topologic map. Nodes are created every 20cm as shown by figure.9. The navigation system is a simplified version of [15] and will be briefly described in the following section. Each created node stores an omnidirectional image and its location within the map including the distance to neighbour nodes and angles.

B. Localization

The robot estimates its position in the map using the activation of nodes. Let Act_i be the activity of a node N_i corresponding to the probability that the robot is at a certain location. The computation of probabilities relies on the use of a conjunction between visual observations and odometric data. The Localization system work as follows :

1) Compute the activity of every node of the map.
2) Set the most active nodes as the current position. If no node is found, create a new node corresponding to the new position.
3) Update nodes information if a new node has been added at step (2).

Considering only odometric data, the estimation of the probability that the robot moved from node N_i to N_j knowing that the travelled a distance r_{od} in a direction θ_{od} is given by :

$$P_{od}(ij/od) = E_1 \times E_2$$

with

$$E_1 = exp(\frac{-(\theta_{ij} - \theta_{ij})}{L^2}),\ E_2 = exp(\frac{-(r_{od} - r_{ij})}{M^2})$$

During the experiments $L = 30°$ and $M = 20cm$. The values of θ_{ij} and r_{ij} are stored in the edge A_{ij}, A_{ij} being the arc connecting two nodes N_i and N_j.

The activity of a node N_i at a step t from odometry is then given by :

$$Act_i^{od}(t) = \max_{j \in S_{N_i}} (Act_i(t-1)P_{od}(ij/od))$$

with S_{N_i} the set of nodes connected to N_i.
In a similar manner, visual probability are written :

$$P_{vis}(I_{N_i}/I_{cod}) = exp(\frac{-(I_{cod} - I_{N_i})}{L^2})$$

with I_{cod} the acquired codified image as explained in section II.C, and I_{N_i} the image stored in N_i. The value of L is

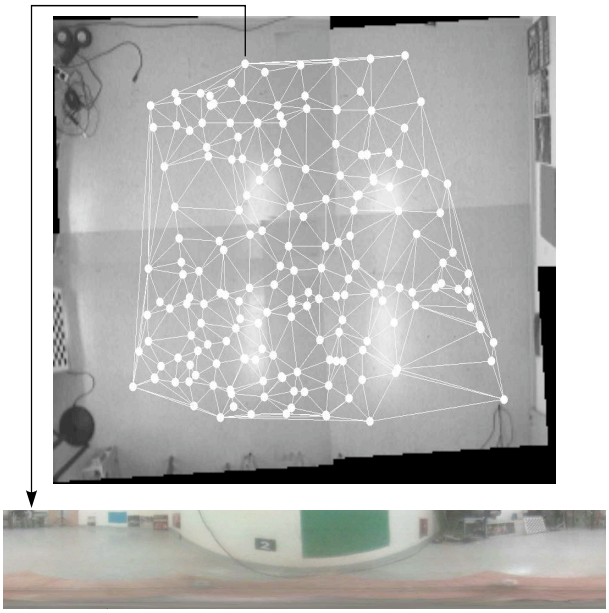

Fig. 9. A topological map showing different marked positions after a random exploration of the environment. Nodes are created every 20cm, they store an omnidirectional image and its location within the map including the distance to neighbour nodes and angles.

connected to the distance between nodes, and should be adjusted so that the activity of a set of nodes connected to the most active one should be 50% of its activity.
The general visual activity of a node N_i at a step t is then given by :

$$Act_i(t+1) = Act_i^{od}(t) P_{vis}(I_{N_i}/I_{cod})$$

The estimated position of the robot is then given by computing the mean position of most active close nodes weighted according to their respective activity.

C. Results

In order to check the robustness of the method, and to study the precision of the Localization of the system series of test are carried out. The acquired images are omnidirectional catadioptric systems, they are transformed into cylindrical images of size 1400×140 pixels. In order to test different scenarios, the content of acquired images panoramic images are modified in the following manner:

- Virtual occlusions are added (from 1 to 10 squares) at random positions, their individual size never exceeding 10% of the size of the original image.
- Independent white noise is added having a uniform distribution between 0 and 255 concerning 10% of the maximum number of pixels.

TABLE I

RESULTS OF LOCALIZATION FOR NONADAPTATIVE PATCHES FOR SEVERAL SIZES.

patch area	Localization rate	computational time
7x7	0.512	1813.27 ms
10x10	0.509487	1452.13 ms
15x15	0.543026	979.4 ms
20x20	0.588974	698.333 ms
25x25	0.611385	553.467 ms
30x30	0.634359	196.933 ms
35x35	0.610872	174.267 ms
40x40	0.589949	158.933 ms

TABLE II

RESULTS FOR THE OPTIMAL PATCH DECOMPOSITION GUIDED BY AN ENTROPY MEASURE.

Index of layer	Localization rate	computation time
1 layer	0.549	55.075 ms
2 layer	0.6991	154.7 ms
3 layer	0.7224	343.5 ms
4 layer	0.7386	588.8 ms
5 layer	0.7451	1 sec et 265 ms
6 layer	0.7503	5 sec et 87 ms

- Additive noise with modification up to 40% of the value of pixels.

to ensure a comparison the optimal patch generation is compared to the uniform decomposition of patches of equal size, with different area values. The presented results were obtained using a mobile $P4M$ $2.20GHz$ and $512Mb$.

Experimental results show that the Localization rates are acceptable even in cases of severe occlusions and noise. The optimal sampling provides a precision of (0.7224) obtained from the third layer. At this stage the mean size of patches is 18 pixels. The fixed sampling has a precision of 0.56 for patches of the same size. The optimal sampling using entropy allows a much important gain. It also two times faster than the fixed sampling. The change of illumination are not properly handled, this is mainly due to the codification of patches rather than the method itself, it is important to notice that the conditions of use $(40\%$ of change) are very important. In normal conditions, the method worked properly for neon light with frequency of $50Hz$ Finaly trajectories

TABLE III

LOCALIZATION RESULTS FOR EACH METHOD IN THE CASE OF NOISE, OCCLUSIONS AND CHANGE IN ILLUMINATION.

method	white noise	illumination	occlusion
Fixed sampling ($30 * 30$)	0.6297	0.5925	0.4996
Optimal patches (6 layers)	0.7358	0.3137	0.6423

are estimated properly, as the mean error is less than the precision of considered nodes.

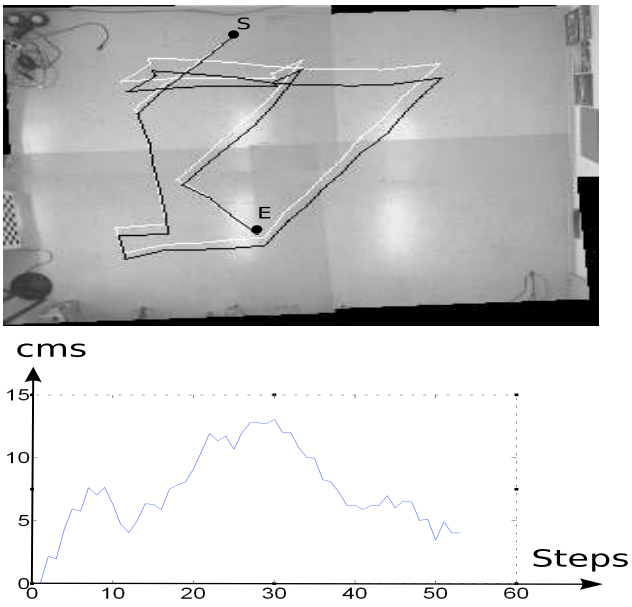

Fig. 10. Estimated trajectory of the robot using the navigation system and ground truth trajectory generated by the camera network.

IV. CONCLUSIONS AND FUTURE WORKS

We have presented a new visual localization method for mobile platforms using bags-of-features. The sampling of adaptive patches techniques shows good results and is efficient in tasks of navigation of robots even in case of noisy and complex environments. It is based on the idea of sampling images using an optimal multilayer quad-tree decomposition of the image driven by entropy. Present work is trying to implement the technique in object recognition applications, where the adaptive sampling could introduce novel recognition approaches.

REFERENCES

[1] Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: ECCV04 workshop on Statistical Learning in Computer Vision. (2004) 5974

[2] Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: CVPR03. (2003) II: 264271

[3] Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. IJCV 43 (2001) 2944

[4] Agarwal, S., Awan, A., Roth, D.: Learning to detect objects in images via a sparse, part-based representation. PAMI 26 (2004) 14751490

[5] Fergus, R., Fei-Fei, L., Perona, P., Zisserman, A.: Learning object categories from googles image search. In: ICCV. (2005) II: 18161823

[6] Grauman, K., Darrell, T.: Efficient image matching with distributions of local invariant features. In: CVPR05. (2005) II: 627634

[7] Leibe, B., Schiele, B.: Interleaved object categorization and segmentation. In: BMVC. (2003)

[8] Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60 (2004) 91110

[9] Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: ECCV. (2002) I: 128

[10] Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: ICCV03. (2003) 14701477

[11] Weber, M., Welling, M., Perona, P.: Unsupervised learning of models for recognition. In: ECCV. (2000) I: 1832

[12] Jurie, F., Triggs, B.: Creating efficient codebooks for visual recognition. In: ICCV. (2005)

[13] Winn, J., Criminisi, A., Minka, T.: Object categorization by learned universal visual dictionary. In: ICCV. (2005)

[14] Agarwal, A., Triggs, B.: Hyperfeatures multilevel local coding for visual recognition. In: ECCV. (2006)

[15] D. Filliat and J. A. Meyer, Map-based navigation in mobile robots I. a review of Localization strategies, Journal of Cognitive Systems Research, vol. 4, no. 4, pp. 243282, 2003.

[16] B. J. Kuipers and Y. T. Byun, A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations, Robotics and Autonomous Systems, vol. 8, pp. 4763, 1991.

[17] D. Hahnel, W. Burgard, D. Fox, and S. Thrun, A highly efficient FastSLAM algorithm for generating cyclic maps of large-scale environments from raw laser range measurements, in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2003

[18] A. Davison, Y. G. Cid, and N. Kita, Real-time 3D SLAM with wideangle vision, in Proc. IFAC Symposium on Intelligent Autonomous Vehicles, Lisbon, July 2004.

[19] A. Tapus and R. Siegwart, Incremental robot mapping with fingerprints of places, in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005), 2005.

[20] E. Krotkov. Mobile robot localization using single image. Proc. *IEEE Int. Conf. on Robotics and Automation*, 978–983, 1989.

[21] I. Horswill, Polly: A vision-based artificial agent, *Proc. of the National Conference on Artificial Intelligence (AAAI)*, 824–829, 1993.

[22] R. Basri and E. Rivlin, Localization and homing using combinations of model views, *Artificial Intelligence*, vol. 78, no. 1-2, pp. 327–354, 1995.

[23] G. Dudek and C. Zhang, Vision-based robot localization without explicit object models, *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 76–82, 1996.

[24] D. Kortenkamp and T. Weymouth, Topological mapping for mobile robots using a combination of sonar and vision sensing, *Proc. of the National Conference on Arti cial Intelligence (AAAI)*, pp. 1972–1978, 1994.

[25] R. Sim and G. Dudek, Learning visual landmarks for pose estimation, *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 1999.

[26] L. Paletta, S. Frintrop, and J. Hertzberg, Robust localization using context in omnidirectional imaging, *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, pp. 2072–2077, 2001.

[27] N. Winters, J. Gaspar, G. Lacey, and J. Santos-Victor, Omni-directional vision for robot navigation, *Proc. IEEE Workshop on Omnidirectional Vision, South Carolina*, 2000.

[28] Z. Dodds and G. Hager, A color interest operator for landmark-based navigation, *Proc. of the National Conference on Artificial Intelligence (AAAI)*, pp. 55–660,1997.

[29] S. Se, D. Lowe, and J. Little, Vision-based mobile robot localization and mapping using scale-invariant features, *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2001.

[30] C. Olson, Probabilistic self-localization for mobile robots, *IEEE Transactions on Robotics and Automation*, vol. 16, no. 1, pp. 55–66, 2000.

[31] B. Krose and R. Bunschoten, Probabilistic localization by appearance models and active vision, *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 1999, pp. 2255–2260.

[32] I. Ulrich and I. Nourbakhsh, Appearance-based place recognition for topological localization, *Proc. of the IEEE International Conference on Robotics & Automation (ICRA)*, 2000, pp. 1023–1029.

[33] F. Dellaert, W. Burgard, D. Fox, and S. Thrun, Using the condensation algorithm for robust, vision-based mobile robot localization, *Proc. of the International Conference on Computer Vision and Pattern Recognition(CVPR)*, 1999.

[34] S. Thrun, Bayesian landmark learning for mobile robot localization, *Machine Learning*, vol. 33, no. 1, pp. 41–76, 1998.

[35] W. Burgard, A. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, Experiences with an interactive museum tourguide robot, *Artificial Intelligence*, vol. 114, no. 1-2, pp. 3–55, 2000.

[36] Osada, R., Funkhouser, T., Chazelle, B., and Dobkin, D. (2002). Shape distributions.

[37] Se, S., Lowe, D., and Little, J. (2001). Vision-based mobile robot localization and mapping using scale-invariant features. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pages 2051–2058, Seoul, Korea.

[38] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features.