2008 IEEE International Conference on
Robotics and Automation
Pasadena, CA, USA, May 19-23, 2008

# Mobile Robot Localization using Panoramic Vision and Combinations of Feature Region Detectors

Arnau Ramisa, Adriana Tapus, *Member, IEEE*, Ramón López de Mántaras, and Ricardo Toledo

***Abstract*— This paper presents a vision-based approach for mobile robot localization. The environmental model is topological. The new approach uses a constellation of different types of affine covariant regions to characterize a place. This type of representation permits a reliable and distinctive environment modeling. The performance of the proposed approach is evaluated using a database of panoramic images from different rooms. Additionally, we compare different combinations of complementary feature region detectors to find the one that achieves the best results. Our experimental results show promising results for this new localization method. Additionally, similarly to what happens with single detectors, different combinations exhibit different strengths and weaknesses depending on the situation, suggesting that a context-aware method to combine the different detectors would improve the localization results.**

***Index Terms*— Affine Regions Detectors, Harris Affine, Hessian Affine, MSER, SIFT, GLOH, Topological Localization**

## I. INTRODUCTION

Finding an efficient solution to the robot localization problem will have a tremendous impact on the manner in which robots are integrated into our daily lives. Most tasks for which robots are well suited demand a high degree of robustness in their localizing capabilities before they are actually applied in real-life scenarios (e.g., assistive tasks).

Since localization is a fundamental problem in mobile robotics, many methods have been developed and discussed in the literature. These approaches can be broadly classified into three major types: metric, topological and hybrid. Metric approaches ([1], [2], [3]) are useful when it is necessary for the robot to know its location accurately in terms of metric coordinates (i.e. Cartesian coordinates). However, the state of the robot can also be represented in a more qualitative manner, by using a topological map (i.e. adjacency graph representation) ([4], [5], [6]). Because the odometry does not provide enough and complete data in order to localize a mobile autonomous robot, laser range finders and/or vision sensors are usually used to provide richer scene information. Furthermore, vision units are cheaper, smaller and more practical than large expensive laser scanners. Therefore, in this work, we propose a topological vision-based localization approach.

In recent years, many appearance-based localization methods have been proposed [7], [8], [9]. SIFT (Scale Invariant Feature Transform) features [9] have been widely used for robot localization. The SIFT approach detects and extracts feature region descriptors that are invariant to illumination changes, image noise, rotation and scaling. Se et al. in [9] used SIFT scale and orientation constraints so as to match stereo images; least-square procedure was used to obtain better localization results. The model designed by Andreasson et al. [10] combines SIFT algorithm for image matching and Monte-Carlo localization; their approach takes the properties of panoramic images into consideration. The work by [11] uses visual landmarks (SIFT features) and geometrical constraints to perform localization.

Another interesting subset of invariant features are the affine covariant regions which can be correctly detected in a wide range of acquisition conditions [12]. Therefore, Silpa-Anan and Hartley in [13] construct an image map based on Harris Affine feature Regions with SIFT descriptors that is later used for robot localization.

The work proposed by Tapus in [5] is closely related to this work. Tapus et al. defined fingerprints of places as generic descriptors of environment locations. Fingerprints of places are circular lists of features and they are represented as a sequence of characters where each character is an instance of a specific feature type. The authors used a multi-perceptional system and global low-level features (i.e., vertical edges, color blobs, and corners) are employed for localization. Nonetheless, our current approach has significant differences from their methodology.

Our novel localization approach uses only panoramic visual information. The *signature of a location* consists of a constellation of feature regions extracted from a panoramic image at a specific location. We decided to use combinations of the following three feature region detectors: the MSER (Maximally Stable Extremal Regions)[14], the Harris-Affine [15], and the Hessian-Affine [12], which have shown to perform better when compared to other region detectors. When a new *signature* is acquired, it is compared to the stored panoramas from the a priori map. The panorama with the highest number of matches is selected as the correspondent. To improve the results and discard false matches, the essential matrix is computed and the outliers filtered. Finally,

This work was partially supported by USC Women in Science and Engineering (WiSE), the FI grant from the Generalitat de Catalunya, the European Social Fund, and the MID-CBR project grant TIN2006-15140-C03-01 and FEDER funds and the grant 2005-SGR-00093.

A. Ramisa is with the IIIA (Artificial Inteligence Research Institute) of the CSIC, UAB Campus, 08193 Bellaterra, Spain. e-mail: aramisa@iiia.csic.es

Dr. Adriana Tapus is with the Robotics Research Lab/Interaction Lab, Department of Computer Science, University of Southern California, Los Angeles, USA. e-mail: tapus@usc.edu

Prof. R. López de Mántaras is with the IIIA (Artificial Inteligence Research Institute) of the CSIC, UAB Campus, 08193 Bellaterra, Spain. e-mail: mantaras@iiia.csic.es

Prof. Ricardo Toledo is with the CVC (Computer Vision Center), UAB Campus, 08193 Bellaterra, Spain. e-mail: ricardo.toledo@cvc.uab.es

the panorama with the highest number of inliers is selected as the best match.

In our approach images are acquired using a rotating conventional perspective camera. When a set of images covering the $360^\circ$ is acquired, they are projected to cylindrical coordinates and the feature regions are extracted and described. The descriptors constellation is next constructed automatically.

Hence, by using feature regions to construct the *signature of a location*, our methodology is much more robust to occlusions and partial changes in the image than the approaches using global descriptors. This robustness is obtained because many individual regions are used for every *signature of a location* and, thus, if some of them disappear the constellation can still be recognized.

This paper is organized as follows. Section II briefly describes the different affine covariant region detectors and descriptors that we used in our work. Section III presents the localization procedure in detail. Experimental results obtained with our mobile robot equipped with a Sony DFW-VL500 camera mounted on a Directed Perception pan tilt unit are presented in Section IV. Finally, Section V contains a discussion of the proposed approach and future research directions.

## II. FEATURE REGIONS AND DESCRIPTORS

An essential part of our approach is the extraction of discriminative information from a panoramic image so it can be recognized later under different viewing conditions. This information is extracted using affine covariant region detectors. These detectors find regions in the image that can be identified even under severe changes in the point of view, illumination, and/or noise.

Recently Mikolajczyk et al. [12] reviewed the state of the art of affine covariant region detectors individually. In this review they concluded that using several region detectors at the same time could increase the number of matches and thus improve the results. Hence, in our work, we have used all the combinations of the following three affine covariant region detectors: (1) Harris-Affine, (2) Hessian-Affine, and (3) MSER (Maximally Stable Extremal Regions), so as to increase the number of detected features and thus of potential matches. Examples of detected regions for the three region detectors can be seen in Fig. 1. These three region detectors have a good repeatability rate, a reasonable computational cost and they are briefly detailed below.

1) The Harris-Affine detector is an improvement of the widely used Harris corner detector. It first detects Harris corners in the scale-space with automatic scale selection using the approach proposed by Lindeberg in [15], and then estimates an elliptical affine covariant region around the detected Harris corners. The Harris corner detector finds corners in the image using the description of the gradient distribution in a local neighbourhood provided by the second moment matrix:

$$M = \begin{bmatrix} I_x^2(x,\sigma) & I_xI_y(x,\sigma) \\ I_xI_y(x,\sigma) & I_y^2(x,\sigma) \end{bmatrix}, \quad (1)$$

where $I(x,\sigma)$ is the derivative at position $x$ of the image smoothed with a Gaussian kernel of scale $\sigma$. From this matrix, the cornerness of a point can be computed using the following equation:

$$R = \text{Det}(M) - k\text{Tr}(M)^2, \quad (2)$$

where $k$ is a parameter usually set to $0.4$. Local maxima of this function is found across the scales, and the approach proposed by Lindeberg is used to select the characteristic scales.
Next, the parameters of an elliptical region are estimated minimizing the difference between the eigenvalues of the second order moment matrix of the selected region. This iterative procedure finds an isotropic region, which is covariant under affine transformations. The isotropy of the region is measured using the eigenvalue ratio of the second moment matrix:

$$Q = \frac{\lambda_{min}(\mu)}{\lambda_{max}(\mu)} \quad (3)$$

where $Q$ varies from 1 for a perfect isotropic structure to 0, and $\lambda_{min}(\mu)$ and $\lambda_{max}(\mu)$ are the two eigenvalues of the second moment matrix of the selected region at the appropriate scale. For a detailed description of this algorithm, the interested reader is referred to [16].
2) The Hessian-Affine detector is similar to the Harris-Affine, but the detected regions are blobs instead of corners. The base points are detected in scale-space as the local maxima of the determinant of the Hessian matrix:

$$H = \begin{bmatrix} I_{xx}(x,\sigma) & I_{xy}(x,\sigma) \\ I_{xy}(x,\sigma) & I_{yy}(x,\sigma) \end{bmatrix}, \quad (4)$$

where $I_{xx}$ is the second derivative at position $x$ of the image smoothed with a Gaussian kernel of scale $\sigma$. The remainder of the procedure is the same as the Harris-Affine: base points are selected at their characteristic scales with the method by Lindeberg and the affine shape of the region if found.
3) The Maximally Stable Extremal Regions (MSER) detector proposed by Matas et al. [14] detects connected components where the intensity of the pixels is several levels higher or lower than the intensity of all the neighboring pixels of the region. Regions selected with this procedure may have an irregular shape, so the detected regions are approximated by an ellipse.

Because affine covariant regions must be compared, a common representation is necessary. Therefore all the regions detected with any method are normalized by mapping the detected elliptical area to a circle of a certain size.

Once the affine covariant regions are detected and normalized, to reduce even more the effects caused by changes in the viewing conditions, these regions are characterized using a feature region descriptor. In our work, we have used Scale Invariant Feature Transform (SIFT) [17] and Gradient Location-Orientation Histogram (GLOH) [18]. These two descriptors were found to be the best in a comparison of

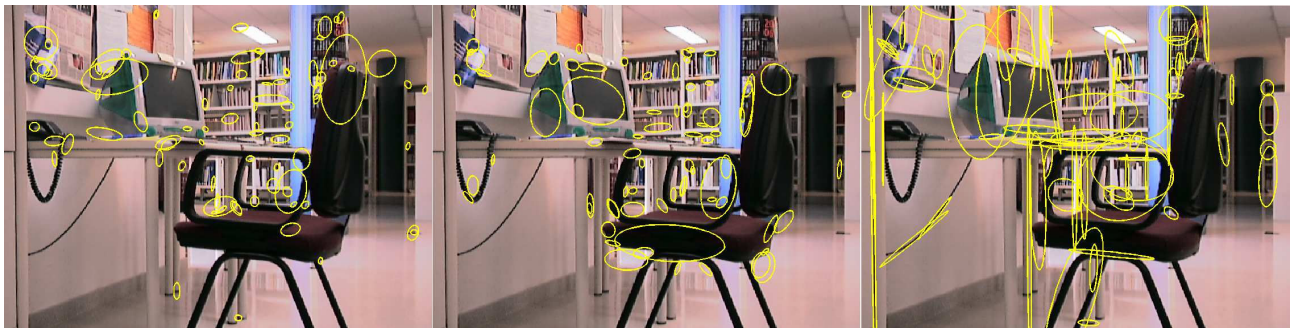

Fig. 1. Example of regions for the three affine covariant region detectors, from left to right: Harris-Affine, Hessian-Affine and MSER.

various state of the art region descriptors [18]. The SIFT descriptor computes a 128 dimensional descriptor vector with the gradient orientations of a detected region. In short, to construct the descriptor vector the SIFT procedure divides the region in 16 rectangular sub-regions and then, for every sub-region, it builds a histogram of 8 bins with the gradient orientations weighted with the gradient magnitude to suppress the flat areas with unstable orientations. The descriptor vector is obtained by concatenating the histograms for every sub-region. The GLOH descriptor is similar to SIFT, with two main differences: the sub-regions are defined in a log-polar way, and the resulting descriptor vector has 272 dimensions but it is later reduced to 128 with a PCA.

These two descriptors are based on the same principle but with slightly different approaches. As they have no complementary properties, our objective in this comparison is to determine which one achieves the best performance. Therefore we have not used them at the same time.

## III. APPEARANCE-BASED LOCALIZATION

The topological localization schema we propose consists in a map represented as a graph where nodes are places visited by the robot, and edges stand for the accessibility information between them. Each node of the graph has an associated *signature*, which, in our case, is a constellation of affine covariant regions characterized with a feature descriptor.

When a novel panoramic image is acquired, a new constellation of features is extracted with the methods described in the previous section, and it is compared with those stored in the map. Finally, the most similar is selected as the corresponding one. The procedure is depicted in Fig. 2. In order to find correspondences between the feature regions of different views, a matching stage is necessary. In this stage each descriptor from the novel constellation is compared to all the descriptors of the other constellation using the Euclidean distance, and the nearest neighbor is selected as the corresponding one. To reject false matches, the distance of the first and the second nearest neighbor are compared, and if they are too similar the match is discarded. The threshold value used to reject false matches is the one proposed by Lowe in [17]:

$$\frac{NN_2}{NN_1} > 0.8, \tag{5}$$

where $NN_1$ is the distance to the first nearest neighbor (the selected as match) and $NN_2$ is the distance to the second nearest neighbor. Lowe found in his work that this distance ratio eliminated 90% of false matches while removing only 5% of correct matches.

The essential matrix [19] is computed using these matches to enforce the geometrical constraints that relate the two views and reject the false correspondences that may have passed the previous stage.

The computation of this matrix is a model fitting process that gives as output both the model itself (the essential matrix) and a subset of correspondeces that agree with the computed model. The bigger the inliers subset, the more similar the novel constellation and the map node. The method used to compute the essential matrix from the found correspondences is the 8-point algorithm with the RANdom SAmple Consensus or RANSAC to reject false matches.

The matchings are classified as inliers and outliers depending on the distance of the points to the epipolar sinusoid described by the essential matrix. As well as in conventional cameras, in cylindrical coordinates the essential matrix verifies:

$$p_0^\top E p_1 = 0, \tag{6}$$

where $p_0$ and $p_1$ are projections of a scene point in the panoramic images, and $E$ is the essential matrix relating the two panoramas. However, contrarily to the case of conventional cameras, the intersection of the projection plane with a cylindrical surface does not define a line but an ellipse, and once the cylinder is unrolled, it appears as a sinusoid. The equation of this sinusoid is:

$$z_1(\phi) = -\frac{n_x cos(\phi) + n_y sin(\phi)}{n_z}, \tag{7}$$

where $z_1(\phi)$ is the height corresponding to the angle $\phi$ in the panorama, and $n_1 = [n_x, n_y, n_z]$ is the epipolar plane normal, obtained with the following expression,

$$n_1 = p_0^\top E. \tag{8}$$

An advantage of the proposed method is that, even though in this work it is conceived as a topologic localization method, it implicitly recovers the essential matrix between the actual view and the reference view. In [20] the authors perform different experiments to assess the accuracy of the computed essential matrix against ground truth data. This essential

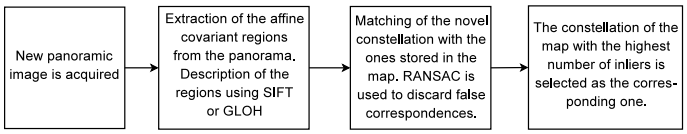


Fig. 2. Steps for panorama-based localization.

matrices can be used to compute the metric localization in reference to the map node using for example the technique proposed in [21]. This information can be then used for metric local navigation with no extra computational cost.

## IV. EXPERIMENTAL RESULTS

The objective of the present work is twofold: In the first place we want to validate the proposed method for global localization and, in second place, we want to experimentally determine if using at the same time different region detectors improves significantly the localization results. Therefore, we acquired multiple panoramas of different rooms and selected some of them as map nodes. Then we used the remaining panoramas to perform a localization test as explained in Section III. Although successive images acquired by the robot while moving in the room could be used to incrementally refine the localization, in this experiment we have only considered the worst case scenario, where only one image per room is available to localize the robot.

The test-bed data used in this work consists in 18 sequences of panoramas from rooms in various buildings[1]. Each sequence consists of several panoramas acquired every 20 cm following a straight line predefined path. This type of sequences are useful to check the maximum distance at which a correct localization can be performed. In order to make the data set as general as possible, rooms with a wide range of characteristics have been selected. For example some sequences correspond to long and narrow corridors, while others have been taken in big hallways, large laboratories with repetitive patterns or individual offices.

The panoramas have been constructed by stitching together multiple views taken from a fixed optical center with a Directed Perception PTU-46-70 pan-tilt unit and a Sony DFW-VL500 camera. The camera and pan-tilt unit can be seen in Fig. 3.

The region detectors and descriptors provided by the authors of [12] at `http://www.robots.ox.ac.uk/~vgg/research/affine/` were used to extract the affine-covariant regions from the images. To construct the panoramas, the images acquired with the camera are projected to cylindrical coordinates, and then the displacement between each pair of images is computed. To compute the displacements, the same feature points used for localization are used and, if not enough points are detected, a correlation-based approach is employed. This approach finds the displacement where the highest correlation between the edges extracted from the images is achieved. The correlation-based method works well even in the case of very low texture, but is more computationally expensive than using the feature matches. Although the panoramic images were constructed for validation purposes, the constellations of feature region descriptors were not extracted from them. Instead, the regions from the original images projected to cylindrical coordinates where used. The reason for this is to avoid false regions introduced by possible new artifacts created during the stitching process. The panoramas built with the stitching method where all correctly constructed, with only some small vertical misalignments, even in the case of changes in lightning, reflections, multiple instances of objects or lack of texture. The sequences have been acquired in uncontrolled environments, with nuisances such as severe illumination changes, repetitive patterns and areas without texture in addition to the changes in point of view.

[1]The data-set can be downloaded from `http://www.iiia.csic.es/~aramisa`.

Fig. 3. The camera and pan-tilt unit used to take the images.

TABLE I

AVERAGE PERCENTAGE OF CORRECTLY LOCALIZED PANORAMAS ACROSS ALL SEQUENCES. FOR CONVENIENCE WE HAVE LABELED M: MSER, HA: HARRIS-AFFINE, HE: HESSIAN-AFFINE, S: SIFT, G: GLOH.

| Combination | Correct Localization |
|---|---|
| M+G | 58.95% |
| M+S | 60.42% |
| HA+G | 68.76% |
| HA+S | 73.55% |
| HE+G | 62.11% |
| HE+S | 58.04% |
| M+HE+G | 59.51% |
| M+HE+S | 57.44% |
| HA+HE+G | 67.18% |
| HA+HE+S | 64.24% |
| M+HA+G | 69.05% |
| M+HA+S | 64.18% |
| M+HA+HE+G | 64.93% |
| M+HA+HE+S | 62.1% |

In order to fulfill our two objectives, we tested all possible combinations of the three selected region detectors with two different descriptors. As can be seen in the Table I, which shows the results of the localization test for every combination, most combinations have an average performance greater than 60% of correct localization across all sequences. The combinations that achieved the best performance in the localization test where Harris-Affine with SIFT and with

TABLE II

AVERAGE PERCENTAGE OF CORRECTLY LOCALIZED PANORAMAS FOR SOME INTERESTING SEQUENCES. THE NAMING CONVENTION IS THE SAME AS IN TABLE I.

| Combination | Lab | Corridor1 | Corridor2 | Conf. Room |
|---|---|---|---|---|
| M+G | 80% | 21% | 15% | 100% |
| M+S | 90% | 11% | 30% | 100% |
| HA+G | 60% | 53% | 25% | 100% |
| HA+S | 30% | 68% | 25% | 100% |
| HE+G | 30% | 84% | 15% | 85% |
| HE+S | 20% | 79% | 10% | 62% |
| M+HE+G | 30% | 21% | 20% | 77% |
| M+HE+S | 10% | 16% | 30% | 54% |
| HA+HE+G | 50% | 89% | 50% | 69% |
| HA+HE+S | 40% | 79% | 55% | 69% |
| M+HA+G | 20% | 21% | 40% | 100% |
| M+HA+S | 70% | 26% | 45% | 69% |
| M+HA+HE+G | 40% | 26% | 35% | 85% |
| M+HA+HE+S | 50% | 32% | 35% | 77% |

GLOH, MSER and Harris-Affine described using GLOH, and Harris-Affine and Hessian-Affine described with GLOH. These methods correctly classify more than a 67% of the test panoramas.

The different region detectors achieved varying results depending on the characteristics of each room. For example those methods that include Hessian-Affine but not MSER performed particularly well in narrow and long corridors. On the other hand, in scenes with numerous repetitive patterns, MSER outperformed the other methods. Table II presents results for some particularly interesting sequences. The first sequence (column "Lab" of Table II) is from the IIIA laboratory, which has a considerable number of repetitive textures due to the barcodes of some artificial landmarks. As can be seen, in this sequence the best performance is achieved by MSER. Another important factor for the superiority of MSER in this sequence is that it is not very long, just about two meters. The second sequence ("Corridor 1") is from a long and narrow corridor of the IIIA. In this sequence the best performance is achieved by the combination of Hessian-Affine and Harris-Affine with the GLOH descriptor, and closely followed by the Hessian-Affine alone. The third sequence ("Corridor 2") is from another corridor, but in this case one of the walls is made out of glass and therefore the exterior can be seen. However, in this sequence bright sunlight has burned the images and only some texture remains. As can be seen in the table, due to the lack of texture, the results for individual methods are very low, but the combinations of different methods (especially Harris-Affine and Hessian-Affine) increase the performance quite a bit.

Finally, the fourth sequence ("Conf. Room") is from the conference room of the IIIA. In this room individual methods had a very good performance, better than the combinations. The conference room has many repetitive textures and a considerable amount of texture, and therefore combinations of different methods have a higher outlier ratio than the cases of just one detector. Another interesting result obtained in this work is the maximum distance at whitch a reliable recognition is possible. This information is useful, for example, to avoid building a too sparse or too dense topological map.

As can be seen in Fig. 4, up to approximately 2.5 meters away from the original point the probability of recognizing a panorama is quite high for all the combinations of methods that achieved the best performance (i.e. Harris-Affine with both GLOH and SIFT; Harris-Affine, Hessian-Affine and GLOH; and MSER, Harris-Affine and GLOH). To compare the results of the chosen detectors and descriptors to another state-of-the-art feature region detector, we performed the same experiments using the method proposed by Lowe in [17]. This method uses as initial points the local maxima of the Differences of Gaussians (DoG), defines a circular region around these initial points, and then SIFT is used to describe the selected regions. For our tests we used the demo program provided by Lowe at `http://www.cs.ubc.ca/~lowe/keypoints/`.

On average, using points detected with the DoG and SIFT, the correct location was selected 51.87% of the times. However, the results were pretty irregular depending on the room. For example, the results from the corridor 1 sequence had only 5% of the panoramas correctly localized, while the conference room of the research center achieved 85% of correct classifications. In most of the sequences, all the affine-covariant region detectors outperformed the results of this detector.

In terms of computational complexity, the current implementation of the method implies comparing all the descriptors from all the panoramas with the descriptors of the new panorama and performing a RANSAC step for every panorama in the database. In order to use this localization method in a real robot, techniques to alleviate this

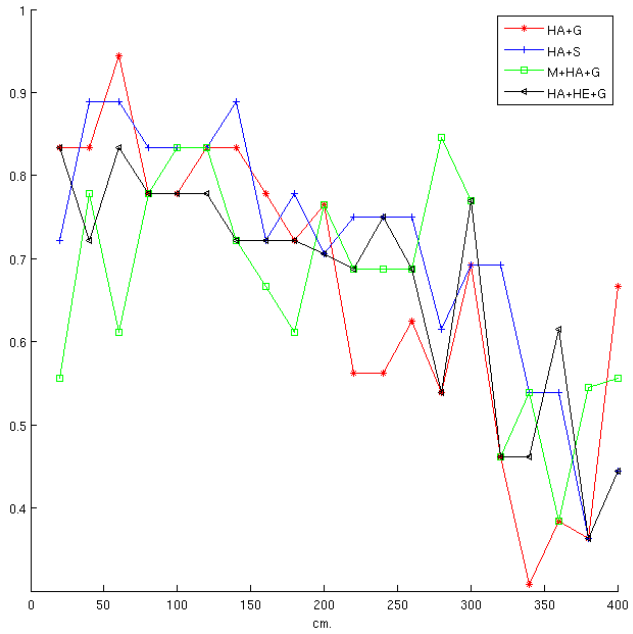


Fig. 4. Percentage of correct localizations against distance for the four better combinations according to Table I. The notation is the same as in the Table.

computational load should be used. Global descriptors to reject unlikely panoramas could greately reduce the number of nodes from the map that must be considered. Another option could be using a K-D tree to accelerate the matching procedure in a similar way as it is done in [17].

## V. CONCLUSIONS AND FUTURE WORK

In this work we proposed and evaluated a *signature* to characterize places that can be used for global localization. This *signature* consists of a constellation of feature region descriptors, computed from affine-covariant regions extracted from a panoramic image acquired in the place we want to add to the map. Later, these *signatures* are compared to the constellation extracted from a new panoramic image using geometric constraints, and the most similar *signature* is selected as the current location. To compare the different *signatures*, the 8-point algorithm with RANSAC to reject false matches is used.

Regarding the validation of the global localization schema, the results obtained show that by using the presented method, a room can be reliably recognized from a distance between two or three meters away from the point where the initial panorama was acquired. The highest score was achieved by the combination of Harris-Affine and SIFT, with which approximately 74% of the localization tests were successful.

We have also compared the results of the proposed affine-covariant region detectors with the scale-invariant region detector proposed by Lowe in [17], widely used in robot navigation, and showed that the affine-covariant regions outperformed Lowe's scale-invariant method.

Different region detectors exhibit different strengths and weaknesses. No single detector had a perfect performance in every situation: Harris-Affine worked well almost everywhere, but in rooms with many repetitive patterns the performance decreased and MSER achieved a higher percentage of success. In narrow and long corridors Hessian-Affine outperformed the other methods. Additionally, tests performed combining different region detectors show that simply using at the same time different types of features does not improve the results directly; nor in the case of different rooms, neither in the maximum distance. However, more advanced methods to combine types of features, such as voting schemas, have shown improvements in similar schemes [22], and perhaps it could also improve our approach.

An interesting line of continuation for this research would be investigating context-aware methods to combine different types of feature regions empowering the strengths of each type while lowering its weaknesses. Another line of continuation that could significantly ameliorate the results would be improving the descriptor matching strategy used, which was not the focus of this work.

## REFERENCES


[1] M. Dissanayake, P. Newman, M., S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, 2001.

[2] A. Castellanos, J. and D. Tardos, J., *Mobile Robot Localization and Map Building: Multisensor Fusion Approach*. Kluwer Academic Publisher, 1999.

[3] S. Thrun, "Probabilistic algorithms in robotics," *Artificial Intelligence Magazine*, vol. 21, pp. 93–109, 2000.

[4] H. Choset and K. Nagatani, "Topological simultaneous localization and mapping (SLAM): Toward exact localization without explicit localization," *IEEE Transactions On Robotics and Automation*, vol. 17, no. 2, pp. 125–137, 2001.

[5] A. Tapus and R. Siegwart, "A cognitive modeling of space using fingerprints of places for mobile robot navigation," in *In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'06)*, (Orlando, USA), pp. 1188–1193, May 2006.

[6] P. Beeson, K. Jong, N., and B. Kuipers, "Towards autonomous topological place detection using the extended voronoi graph," in *IEEE International Conference on Robotics and Automaton (ICRA)*, (Barcelona, Spain), pp. 4373–4379, 2005.

[7] C. Owen and U. Nehmzow, "Landmark-based navigation for a mobile robot," in *From Animals to Animats: Fifth International Conference on Simulation of Adaptive Behavior (SAB)*, (Cambridge, MA), pp. 240–245, MIT Press, 1998.

[8] M. Franz, O. Schlkopf, B. Mallot, and A. Blthoff, H., "Learning view graphs for robot navigation," *Autonomous Robots*, vol. 5, pp. 111–125, 1998.

[9] S. Se, D. Lowe, and J. Little., "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *International Journal of Robotics Research (IJRR)*, vol. 21, no. 8, pp. 735–758, 2002.

[10] H. Andreasson, A. Treptow, and T. Duckett, "Localization for mobile robots using panoramic vision, local features and particle filters," in *In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'05)*, (Barcelona, Spain), 2005.

[11] O. Booij, Z. Zivkovic, and B. Krose, "From sensors to rooms," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) Workshop - From Sensors to Human Spatial Concepts*, (Beijing, China), pp. 53–58, 2006.

[12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, and L. V. Kadir, T.and Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 2, pp. 43–72, 2005.

[13] C. Silpa-Anan and R. Hartley, A., "Localization using an image-map," in *In Proceedings of the 2004 Australasian Conference on Robotics and Automation*, (Canberra, Australia), 2004.

[14] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *In Proceedings of the British Machine Vision Conference (BMVC'02)*, (Cardiff, UK), 2002.

[15] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[16] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *Int. J. Comput. Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[17] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[18] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second ed., 2004.

[20] A. Ramisa, R. Lopez de Mantaras, D. Aldavert, and R. Toledo, "Comparing combinations of feature regions for panoramic vslam," in *4th International Conference on Informatics in Control, Automation and Robotics (ICINCO'07)*, (Angers, France), 2007.

[21] S. B. Kang and R. Szeliski, "3-D scene data recovery using omnidirectional multibaseline stereo," in *CVPR '96: Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, (Washington, DC, USA), pp. 364–370, IEEE Computer Society, 1996.

[22] D. Aldavert, A. Ramisa, and R. Toledo, "Wide baseline stereo matching using voting schemas," in *1st CVC Research and Development Workshop, 2006*, pp. 30–36, Computer Vision Center, 2006.