Joan Martí José Miguel Benedí
Ana Maria Mendonça Joan Serrat (Eds.)

LNCS 4478

Pattern Recognition and Image Analysis

Third Iberian Conference, IbPRIA 2007
Girona, Spain, June 2007
Proceedings, Part II

2 Part II

Springer

Lecture Notes in Computer Science 4478

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

Joan Martí José Miguel Benedí
Ana Maria Mendonça Joan Serrat (Eds.)

Pattern Recognition and Image Analysis

Third Iberian Conference, IbPRIA 2007
Girona, Spain, June 6-8, 2007
Proceedings, Part II

Springer

Volume Editors

Joan Martí
University of Girona
Campus Montilivi, s/n., 17071 Girona, Spain
E-mail: joanm@eia.udg.es

José Miguel Benedí
Polytechnical University of Valencia
Camino de Vera, s/n., 46022 Valencia, Spain
E-mail: jbenedi@dsic.upv.es

Ana Maria Mendonça
University of Porto
Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal
E-mail: amendon@fe.up.pt

Joan Serrat
Centre de Visió per Computador-UAB
Campus UAB, 08193 Belaterra, (Cerdanyola), Barcelona, Spain
E-mail: joan.serrat@cvc.uab.es

Library of Congress Control Number: 2007927717

CR Subject Classification (1998): I.4, I.5, I.7, I.2.7, I.2.10

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-72848-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72848-1 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12070374 06/3180 5 4 3 2 1 0

A New Method for Robust and Efficient Occupancy Grid-Map Matching

Jose-Luis Blanco, Javier Gonzalez, and Juan-Antonio Fernandez-Madrigal

Department of System Engineering and Automation
University of Malaga
Malaga 29071, Spain

Abstract. In this paper we propose a new matching method for occupancy grid-maps under the perspective of image registration. Our approach is based on extracting feature descriptors by means of a polar coordinate transformation around highly distinctive points. The proposed method presents a modest computation complexity, although it can find matchings between features reliably and regardless their orientation. Experimental results show the robustness of the estimates even for dynamic environments. Our proposal has important applications into the field of mobile robotics.

1 Introduction

Occupancy grid-maps, introduced into the robotics community two decades ago [1], are a very valuable representation for map building applications of planar environments [2]. In this representation, the space is arranged in a metric grid of cells that store the probability of that area being occupied by some obstacle. A recent trend in map-building research is to consider hierarchical models, where each node within a topological graph represents a local metric map [3]. A critical issue for this paradigm is to detect when two local maps correspond to the same physical place, and, in that case, to compute the relative transformation between those maps. Solving this problem is crucial for the consistency of the mapping process. The aim of the present work is to provide a solution to this problem from an image registration viewpoint when local maps are occupancy grid-maps.

Occupancy grids can be naturally interpreted as grayscale images (called here), where cells in the grid correspond to pixels in the image, thus by registering the images we obtain the spatial transformation between the maps. Image registration techniques can be straightforwardly grouped into intensity-based ones, and those based on feature extraction (see [4] for a review). Although the former approach has been already applied to grid-map matching [2], an approach based on feature extraction, as the one presented here, is less computationally expensive, becoming more appropriate for being integrated into a real-time mapping framework.

Our overall approach consists of the following three steps: (i) feature-point detection in the map images and extraction of their descriptors, (ii) estimation

J. Martí et al. (Eds.): IbPRIA 2007, Part II, LNCS 4478, pp. 194–201, 2007.

of the likely correspondences between features, and (iii) robust estimation of the rigid transformation between the maps. Since the cell size of all the maps can be set to any fixed value, there are not differences in scale in this problem. Taking this into account, in this paper we propose a new descriptor and an associated method for finding correspondences that are able to efficiently and robustly solve correspondences between feature points in map images of arbitrary orientation. Other previously proposed descriptors in the literature, in spite of being very useful for dealing with real images taken from cameras, become unpractical here due to different reasons:

- The Scale Invariant Feature Transform (SIFT) descriptor, introduced in [5], implies much more computation effort than required for the problem addressed here, since it achieves scale invariance by constructing a pyramid of auxiliary sub-sampled images.
- In [6] it is presented a descriptor that, although based on polar coordinate transformation like ours, proposes an additional step for extracting moments from the Fourier transform. However, we have experimentally verified that this method is not as well suited as ours to effectively discriminate between features typically found in map images.
- In [7] it is proposed to take Gaussian derivatives as descriptors, in the context of developing an affine invariant descriptor. We believe that the low dimensionality of the descriptor proposed there is not appropriate for the highly ambiguous features in map images.

In the next section we describe our proposal for a feature point descriptor in map images. Next, section 3 describes the associated methods for measuring the degree of matching between a pair of features and how to robustly estimate the map displacement from those matchings. Finally, in section 4 we provide experimental results for different map matching situations, all of them employing real data.

2 The Cylindrical Descriptor

We assume that a set of N feature points $\varphi = \{\mathbf{p}_1, ..., \mathbf{p}_N\}$ has been extracted from a map image using any appropriate method with a good repeatability. In this work we employ the method proposed by Shi and Tomasi [8], although using other methods, like the Harris corner detector [9], leads to similar results.

Once a feature point $\mathbf{p}_a = [x_a \; y_a]^T$ has been localized, we define its associated descriptor f_a as a mapping of the annular area around the feature point into the two-dimensional space of polar coordinates r and θ (refer to Fig. 1). Notice that the cylindrical topology of this transformed space can be interpreted as a "panoramic image" of the neighborhood of the feature point, as shown with an example in Fig. 1(c)–(d). Hence it is clear that a rotation in the grid-map becomes a rotation of the cylindrical image around the θ axis. Here we consider radial distances only within the range $[R_{min}, R_{max}]$, e.g. from 0.10 to 1.50 meters, and implement the descriptor as a $N_r \times N_\theta$ matrix with dimensions $N_r = (R_{max} - R_{min})/\Delta_r$ and

$N_\theta = 2\pi/\Delta_\theta$, provided the desired spatial and angular resolutions Δr and $\Delta\theta$, respectively. The value of the descriptor for each pair (i, j) in the range $[0, N_r - 1]\times[0, N_\theta - 1]$ is given by integration over the corresponding annular sector (please, refer to Fig. 1(a)–(b)):

$$f_a[i,j] = \int_{\phi_j}^{\phi_{j+1}} \int_{r_i}^{r_{i+1}} m\left(\begin{bmatrix} x_a + r\cos\theta \\ y_a + r\sin\theta \end{bmatrix}\right) dr d\theta \tag{1}$$
$$r_i = R_{min} + i\Delta r$$
$$\phi_j = j\Delta\phi$$

where $m(\mathbf{x})$ represents the contents of the map at the 2D point $\mathbf{x}$. Notice that, in practice, the above integration can be computed through a Monte-Carlo approximation, where a number of points within the integration area are evaluated with sub-pixel precision by straightforward cubic interpolation.

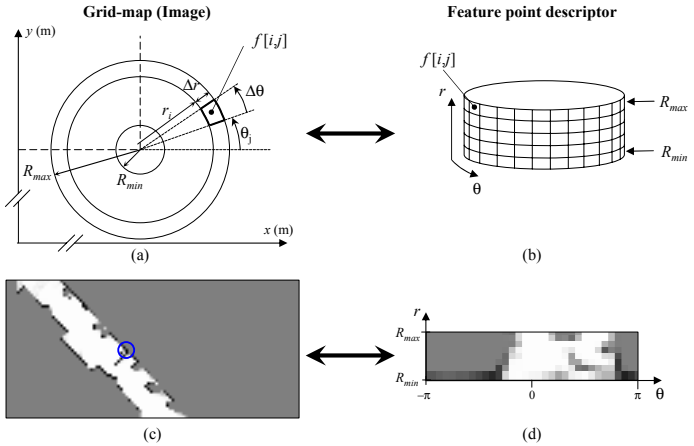

Fig. 1. (a)-(b) The geometry of the descriptor proposed in the text, which maps the circle around the feature into a cylindric space. An example is shown in (c)-(d).

3 Map Matching

3.1 Measuring the Degree of Matching Between a Pair of Descriptors

As a motivating example, please consider the pair of features detected in the maps of Fig. 2(a)–(b), which correspond to the same physical point. The associated descriptors are shown in Fig. 2(c). It is clear that their cylindrical descriptors will be very similar for some shift in θ if the features represent a valid

correspondence. In this particular example that shift is 214°, and the similarity between the conveniently rotated descriptors is patent in Fig. 2(d). Hence we propose to measure the degree of matching $d(f_a, f_b)$ for a pair of descriptors f_a and f_b through the minimum Euclidean distance between the descriptors, taken over all possible rotations:

$$d(f_a, f_b) = \min_{j_0 \in [0, N_\theta - 1]} \sum_{i=0}^{N_r - 1} \sum_{j=0}^{N_\theta - 1} \left(f_a[i,j] - f_b[i, (j - j_0) \bmod N_\theta] \right)^2 \qquad (2)$$

Once a matching measure is defined for pairs of features, it must be addressed how to obtain the whole set of correspondences $\mathcal{C} = \{\mathcal{C}_1, ..., \mathcal{C}_k\}$, where each correspondence $\mathcal{C}_i = \langle a_i, b_i \rangle$ consists of a pair of feature indexes a_i and b_i, one from each map. When (2) is evaluated for a fixed feature in the first map and all the features in the other, we expect to obtain a low distance (a good matching) only for a few (ideally only one) of the possible correspondences. An example is shown in Fig. 2(f), where the correct correspondence is clearly differentiated from the rest of associations. Provided that a robust association step will be applied next, it is not a problem to establish at this point more than one correspondence for each feature, thus the following compatibility test will be sufficient for finding the set $\mathcal{C}$.

Firstly, the matching of f_a with the candidate f_b must be sufficiently differentiated from the rest. This condition can be formulated as the distance $d(f_a, f_b)$ to be below a dynamic threshold $\tau_d = \mu - \kappa\sigma$, where μ and σ are the mean and

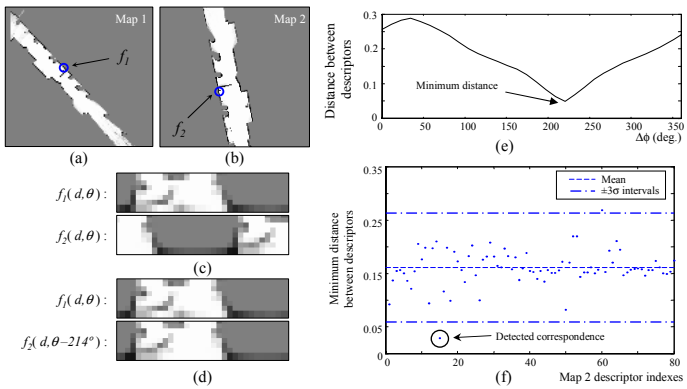

Fig. 2. Two maps of the same environment are shown in (a)–(b), while the descriptors corresponding to the highlighted features are shown in (c) and (d), for a shift in θ of 0^o and 214^o, respectively. The matching distance between those features is plotted in (e) for all the possible rotation angles, and in (f) it is shown the minimum distance between the feature f_1 and all the features in the second map, from where the right correspondence is clearly revealed.

standard deviation, respectively, of the evaluation of $d(f_a, f_j)$ for all the possible values of j. The selectivity of this threshold is controlled by the parameter κ. Any value in the range 1.5-3.0 is appropriate for most situations, although the higher its value, the more demanding we are in accepting a correspondence, at the cost of finding less of them. Secondly, to cope with features without a valid correspondence, we must set a fixed threshold τ_f for the maximum distance between descriptors to be accepted as a correspondence. This parameter, determined heuristically, has been set to 0.07 for all the experiments in this paper. This algorithm is summarized in Table 1.

Table 1. The algorithm for finding compatible correspondences between maps

algorithm `findCorrespondences`$(m_1, m_2) \mapsto \mathcal{C}$
  $\mathcal{C} = \emptyset$
  for each $f_i \in m_1$
    $\mu = \mathrm{E}_j\{d(f_i, f_j)\}$; Mean and standard deviation, where
    $\sigma = \sqrt{\mathrm{E}_j\{(d(f_i, f_j) - \mu)^2\}}$; j spans over all features in m_2
    $\tau_d = \mu - \kappa\sigma$; Compute the dynamic threshold
    for each $f_j \in m_2$
      if $d(f_i, f_j) < \min(\tau_d, \tau_f)$; Compatibility test
        $\mathcal{C} = \mathcal{C} \cup \langle i, j \rangle$; Accept the correspondence
end

3.2 Robust Estimation of the Rigid Transformation Between Maps

Given any set of correspondences, it is well known that a closed-form solution exists for finding the rigid transformation between the maps that is optimal, in the least-minimum-square-error (LMSE) sense [10]. Let this method be denoted by $T(\mathcal{C}_i) \mapsto \mathbf{x_i}$, where $\mathbf{x_i} = [x_i \ y_i \ \phi_i]^T$ is the optimal transformation according to correspondences $\mathcal{C}_i$. However, applying this estimation directly to the whole set of detected correspondences is not convenient, since a wrong correspondence may lead to a large error in the estimated transformation. That is the reason why we propose here an additional RANSAC-based [11] step for robustly estimating the map transformation, what is described in Table 2. In short, we randomly choose a pair of correspondences (the minimum number required), and then all the correspondences that are consistent with the initial estimation are included, providing a robust estimate $\mathbf{x}_i$. Since the choice for the pair of initial correspondences is determinant for the rest of accepted ones, we repeat this process a number of times M, each time with a randomly chosen initial pair of correspondences. Additionally, only those sets of correspondences of a minimum size C_{min} (e.g. 8 correspondences) are considered, achieving improved consistency in the results. In this way, we obtain a set of robust estimates $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{L}$. If we assume the correspondence between features to be an unknown random variable, this set $\mathbf{X}$ can be interpreted as a sample-based (Monte-Carlo) approximation to the probability density of the map transformation, which can be used, for example, for fitting a Gaussian distribution for the maps transformation.

Table 2. The method for robustly estimating the transformation

```
algorithm robustEstimation(C) ↦ X
  X = ∅
  for i = 1..M do                          ; Repeat the simulation M times.
    randomly choose C_i = {c_1, c_2} ⊂ C, such as c_1 ≠ c_2
    x_i = T(C_i)
    for each c_j ∈ C − C_i
      if ‖T(C_i ∪ c_j) − x_i‖ < τ          ; If the new estimation is consistent
        C_i = C_i ∪ c_j                    ;  according to a given threshold τ,
        x_i = T(C_i)                       ;  accept the correspondence c_j.
    if |C_i| ≥ C_min
      X = X ∪ x_i
end
```

4 Experimental Results and Conclusions

We have applied our method to two pairs of maps obtained from real data gathered by a mobile robot in the same physical places, but at different times. As shown in Fig. 3, the pairs of image maps contain some differences, especially the pair in Fig. 3(a) where several pieces of furniture were moved within the room. The computed map transformations are shown in Fig. 3(c)–(f). It is noticeable the high robustness when establishing correspondences, what is reflected in the low uncertainty of the estimations: below 15 cm. for the translation, and less than 2 degrees for the orientation. The estimation process takes 600ms and 807ms for the two pair of maps, respectively, for a number of simulations $M = 5000$. We have also intensively tested the performance of our approach against two kinds of realistic errors that can appear in occupancy grids built from range scans [2]: errors in the ranges themselves, and in the localization of the sensor within the map. Both errors have been simulated by additive Gaussian noise, characterized by σ_{range} and σ_{pose}, respectively. In this experiment we have arbitrarily chosen a map as reference and synthetically generated a test map with a known transformation of $(\Delta x, \Delta y, \Delta\phi) = (1m, 2m, 45^{\circ})$ to compute the mean errors achieved by our method, both in translation ϵ_{XY} and in orientation ϵ_{ϕ}. Errors have been computed for a set of different error levels σ_{range} and σ_{pose}. We have also contrasted our estimation with that from the LMSE method applied on the whole set of correspondences. All these results are summarized in Fig. 4, where it should be highlighted the small absolute errors achieved over the wide range of noise levels and for both kind of errors, in the range values, Fig. 4(d)-(f), and in the poses, Fig. 4(g)-(i). In all the cases the mean errors are below 10 cm. and 0.5 degrees. In comparison with the LMSE estimate, our method achieves an improvement of above one order of magnitude, clearly justifying the integration of the robust step in the process.

We have also computed the estimation based on the normalized cross correlation (NCC) for comparison purposes (see Fig. 4(j)), where it is clear that the maximum value of the NCC reveals the transformation between the maps,

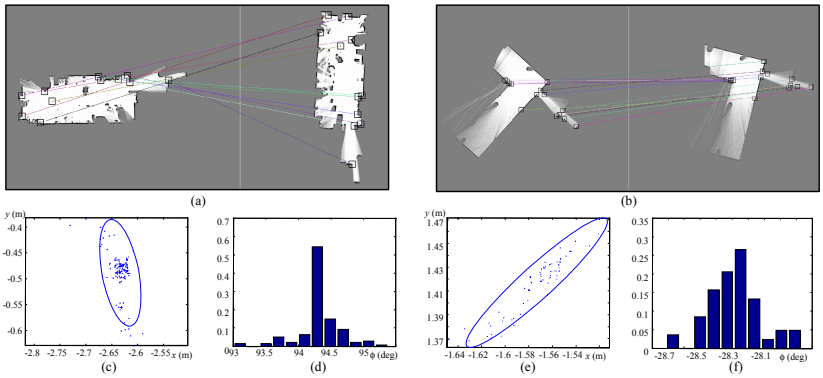

Fig. 3. Matching results from our method for two pairs of real maps, shown in (a)–(b). The samples obtained from the estimation process are shown in (c)–(f), where the estimated translations and orientations have been separated for ease of visualization. Gaussian fit is shown for the translation estimations and a 95% confidence interval.

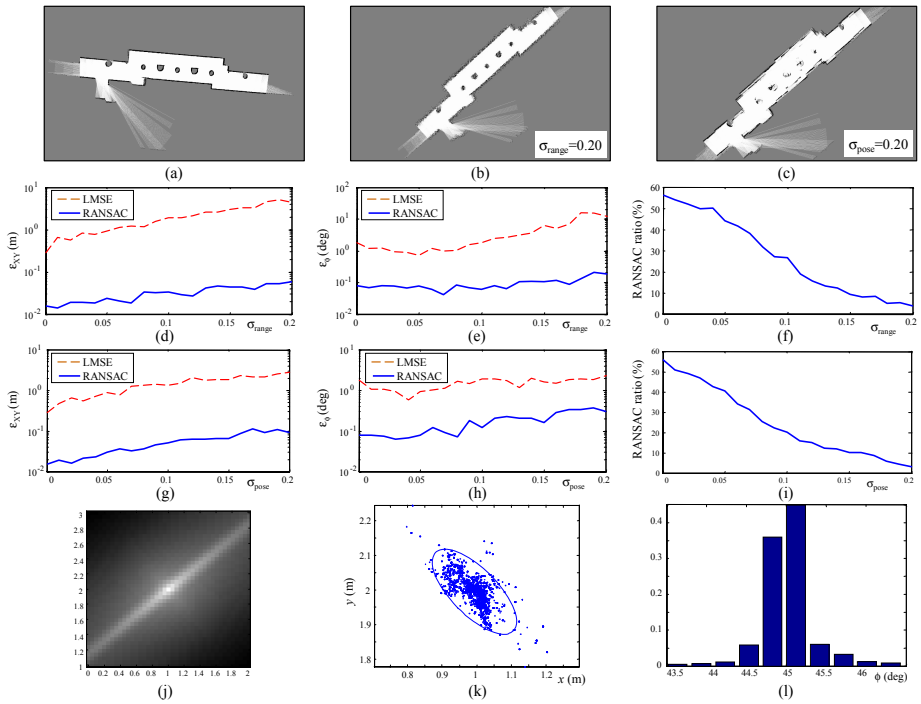

Fig. 4. (a) The reference map, which is displaced and corrupted with noise in sensor measurements (b), and in the sensor localization (c). (d)-(i) Show the performance of our method for different levels of noise. The result from NCC (for a fixed value of $\Delta\phi = 45^o$) is shown in (j), whereas (k)-(l) show the samples obtained from our method, where the fitted Gaussian is represented by its 95% confidence interval. See the text for further details.

but it also assigns high values to many wrong transformations, which contrasts with the results from our approach in Fig. 4(k). Regarding computation time, it takes approximately 420 sec. to evaluate the NCC in a 3.2GHz Pentium 4 using a straightforward implementation, whereas out method takes less than 1 sec.

To summarize, in this paper we have presented a new method for robust matching of occupancy grid-maps, a technique with many potential applications in robotics. Our approach has been devised from a image-registration viewpoint, hence we introduce a new feature-point descriptor for easing the matching. Adding a robust step to the estimation process is shown to provide a significant improvement in the overall precision. Future work should be aimed to provide a more detailed comparison between the performance attainable from different feature-point detectors, and to integrate this work into robotic mapping frameworks.

Acknowledgments. This work was partly supported by the Spanish Government under research contract DPI2005-01391.

References

1. Elfes, A.: Using occupancy grids for mobile robot perception and navigation. Computer 22(6), 46–57 (1989)
2. Gutmann, J., Konolige, K.: Incremental mapping of large cyclic environments. In: Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp. 318–325 (1999)
3. Estrada, C., Neira, J., Tardos, J.: Hierarchical SLAM: Real-Time Accurate Mapping of Large Environments. IEEE Transactions on Robotics 21(4), 588–596 (2005)
4. Zitova, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing 21(11), 977–1000 (2003)
5. Lowe, D.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2 (1999)
6. Matas, J., Bılek, P., Chum, O.: Rotational Invariants for Wide-baseline Stereo. In: Proceedings of the Computer Vision Winter Workshop, vol. 2, pp. 296–305 (2002)
7. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: Proceedings of European Conference on Computer Vision, vol. 1, pp. 128–142. Springer, Heidelberg (2002)
8. Shi, J., Tomasi, C.: Good features to track. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 593–600 (1994)
9. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of Alvey Vision Conference, vol. 15 (1988)
10. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)
11. Fischler, M., Bolles, R.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6), 381–395 (1981)