

Robust Localization Using Panoramic View-Based Recognition *

Matjaž Jogan and Aleš Leonardis

Faculty of Computer and Information Science, University of Ljubljana
Tržaška 25, 1001 Ljubljana, Slovenia
{matjaz.jogan,alesl}@fri.uni-lj.si

Abstract

The results of recent studies on the possibility of spatial localization from panoramic images have shown good prospects for view-based methods. The major advantages of these methods are a wide field-of-view, capability of modelling cluttered environments, and flexibility in the learning phase. The redundant information captured in similar views is efficiently handled by the eigenspace approach. However, the standard approaches are sensitive to noise and occlusion. In this paper, we present a method of view-based localization in a robust framework that solves these problems to a large degree. Experimental results on a large set of real panoramic images demonstrate the effectiveness of the approach and the level of achieved robustness.

1. Introduction and motivation

When dealing with autonomous systems that freely move in space, an important problem to solve is the estimation of the instantaneous position. In the case of autonomous robot navigation, localization is necessary for motion planning. In augmented reality applications, localization of the observer is crucial for registration that allows a combination of virtual and real environments.

In our work we define the problem of localization as the task of recognizing a panoramic view (see Fig. 1 for an example of cylindrical panoramic images) from a set of panoramic views acquired in the learning phase. In the last decade many researchers have shown that feasible models of the world can be constructed without using precise geometrical information [2, 6, 7]. Namely, a model of the world can be constructed as a memory map, built from adequately compressed sets of images. Such methods have been suc-

*This work was supported by the Ministry of Science and Technology of Republic of Slovenia (Project J2-0414 and SI-CZ Intergovernmental S&T Cooperation Programme).

cessfully tested in the areas of object [5, 7] recognition. The main motivation for applying such an approach to the problem of localization is the analogy between recognizing an object in the scene and recognizing the environment. In contrast to object recognition, the target to be recognized in the case of localization is not only a part of the image (on a cluttered background), but rather the complete image. If we use panoramic images as representations of positions, we can expect that views taken from nearby positions and oriented in the same way tend to be strongly correlated as it is in the case of looking at an object from two nearby viewpoints (Fig. 1). This allows us not only to design an efficient strategy based on correlation, but also to build a compact representation that eliminates redundancy.



Figure 1. Two cylindrical panoramic images (labeled 50 and 53 in the path set) taken from viewpoints 60 cm apart.

Another motivation comes from the discoveries on navigation strategies of insects, that are, although limited in the brain size, capable of amazingly confident navigation and of self-localization. In fact, some studies (see [4] and the references therein) imply, that wood ants may use a representation of the environment that is built from wide-angle snapshots of the scene. Localization is then performed by comparing the instantaneous view with the stored snapshots. According to this and some other studies the patterns are processed retinotopically, i.e., the snapshot is not segmented, but interpreted as a whole.

For building a compact model from a set of images, the *eigenspace* approach proved itself as a viable one [3]. A similar work was done by Aihara et al. [1] who used

row-autocorrelated transforms of cylindrical panoramic images in order to achieve invariance to rotation of the sensor around the optical axis. The approach suffers from less accurate results on novel positions, since by correlating the images some of the information is lost. An alternative approach was proposed by Pajdla and Hlaváč [8] who used an appearance-preserving rotational invariant representation, i.e, the *Zero Phase Representation (ZPR)*.

The major limitation of these approaches is the sensitivity of the matching stage to noise and occlusion. It is clear that one has to cope with occlusions in the scene, such as, for example, people walking by, other objects being moved around the environment etc. In this paper, we propose a method for robust localization by applying a robust procedure for recovery of parameters from the eigenspace [5].

The paper is organized as follows. In section 2 we first discuss the major properties of panoramic images and the distribution of their correlation over the sensed environment. In section 3 we describe the procedure for building the environment model from panoramic views and give an overview of the robust recognition of views. In section 4 we present the results on non occluded and occluded data. We conclude with a summary and an outline of future work.

2. Correlation of panoramic images

We have already emphasized the analogy between localization and object recognition. When looking at an object from two nearby viewpoints, there is a high probability that the two views are very similar to each other. If the panoramic sensor has a fixed orientation, as if using an external compass, two images taken at nearby positions also tend to be strongly correlated. As it can be seen in Fig. 2, the distribution of correlation is far from a simply characterized function, however, it gives a good indication of the current location.

Of course, we cannot expect that an external compass is always available. In such a case, one has to employ a transformation that maps cylindrical panoramic images into a representation that is invariant to the rotation of the sensor and also preserves the properties of the correlation distribution. As it was shown in [3], this can be achieved by using a transformation that preserves appearance, such as the ZPR transform, proposed in [8].

3. Panoramic eigenspace

As already stated we represent the environment by a set of panoramic images $\mathcal{I} = \{\mathbf{x}_1 \dots \mathbf{x}_N\}$, taken in the learning phase at arbitrary positions. We transform the images so that they are all oriented the same way. This enables us to efficiently compress them by the eigenspace method.

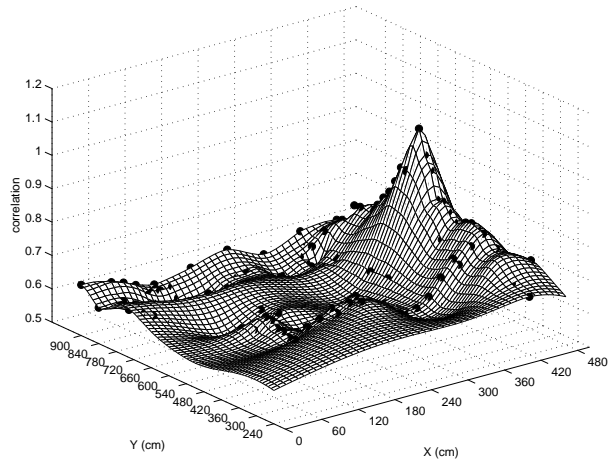


Figure 2. Correlation of panoramic images in space. XY plane represents the coordinates of the experimental environment. Measured and then interpolated is the correlation with image at X=440, Y=660.

The eigenspace method consists of solving the *Singular Value Decomposition* on the covariance matrix of the (normalized) images in \mathcal{I} , to obtain an orthogonal set of vectors $\mathbf{e}_1, \mathbf{e}_2 \dots, \mathbf{e}_n$, usually referred to as *eigenimages*. If we then choose a subset of p eigenimages with the largest *eigenvalues*, we can approximate in the least squares sense each image parametrically as a linear combination of that subset to a desirable degree of accuracy. Namely, every model image \mathbf{x}_i therefore projects into some point \mathbf{q}_i in the *eigenspace*, spanned by the selected eigenimages [7].

The major advantage of the eigenspace method is that the correlation in the image space is related to the Euclidean distance in the eigenspace, i.e., the stronger are the two images correlated, the closer will their projections lie in the eigenspace. It is therefore possible to densely interpolate the set of points to obtain a spline that represents an approximation of an arbitrarily dense set of real-world images [7]. Panoramic views from intermediate positions are in that way approximated by a spline.

3.1. Robust recognition

Once the model is built, recognition of a view is performed by recovering the coefficient vector \mathbf{q} of the instantaneous image \mathbf{y} , or searching for the point on the spline which is the nearest to the projected point. As every point \mathbf{q} is associated with the position parameters, we can make an estimation of the current position. The standard method to recover the parameters is to project the image vector onto

the eigenspace [7]:

$$q_j(\mathbf{y}) = \langle \mathbf{y}, \mathbf{e}_j \rangle; \quad j = 1 \dots p . \quad (1)$$

However, this way of calculation of parameters is non-robust and thus not accurate in the case of noisy or occluded data. If we imagine a mobile robot roaming around with a model acquired under a set of stable conditions, every change in the environment, such as displaced objects, people walking around etc., can result in severe occlusions with respect to the original stored images.

To overcome this problem, we propose to use the robust approach [5], that, instead of using the whole image vectors, generates and evaluates a set of hypotheses \mathbf{r} as subsets of image points $\mathbf{r} = (r_1, r_2, \dots, r_k)$. In fact, the coefficients can be retrieved by solving a set of linear equations on $k = n$ points:

$$x_{r_i} = \sum_{j=1}^n q_j(\mathbf{x}) e_{jr_i} \quad 1 \leq i \leq n . \quad (2)$$

The principle of such computation is illustrated in Fig. 3.

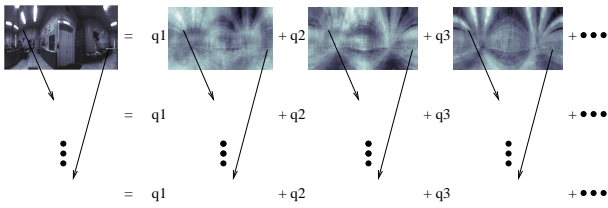


Figure 3. Calculating the coefficients from a set of linear equations.

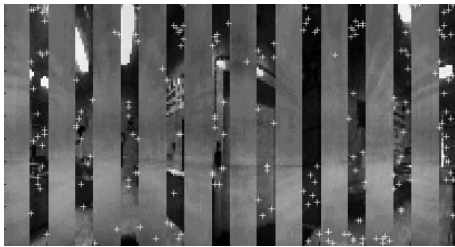


Figure 4. Image at 60% occlusion. Crosses denote the points that contribute to the generation of a hypothesis.

By selecting only p , $p \leq n$ eigenimages as our basis, we cannot use the previous set of equations, but we rather try to solve an over-constrained system in a robust way, so that the solution set of parameters minimizes

$$E(\mathbf{r}) = \sum_{i=1}^k (x_{r_i} - \sum_{j=1}^p q_j(\mathbf{x}) e_{jr_i})^2 . \quad (3)$$

We solve the system on k , $k > p$ points, where k is significantly smaller than the total number of image points. The set of points is randomly selected and due to the robust solving of the equation, only the points on which the error is arbitrary small contribute to the computation of the parameters. As we can see in Fig. 4, at this stage most of the points in the occluded regions are excluded from the computation.

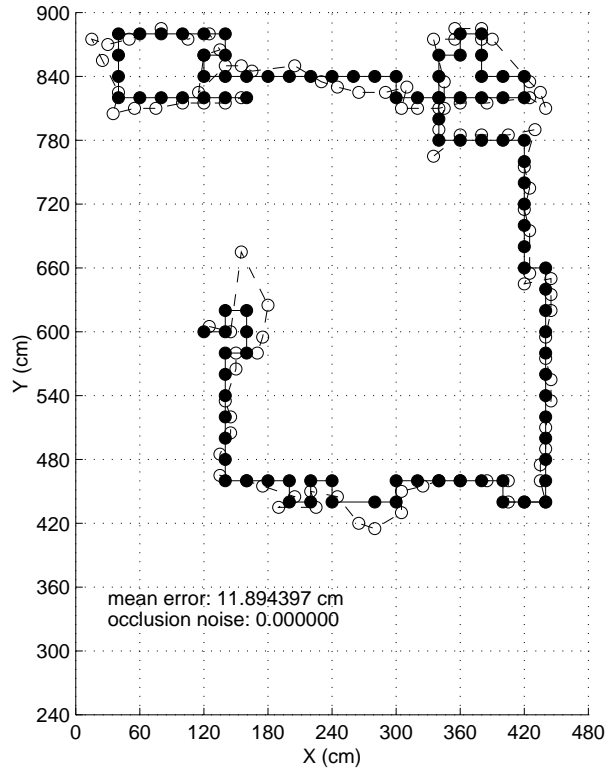


Figure 5. Localization on an imaginary path of 100 images.

To increase the probability of avoiding points that are noise or represent occlusion, several different hypotheses are generated. A hypothesis consists of a set of parameters, an error vector ϵ calculated as the squared difference between the data and the reconstruction, and the domain of compatible points that satisfy an error margin constraint. These hypotheses are then subject to a selection procedure, based on the *Minimal Description Length* principle, as described in [5].

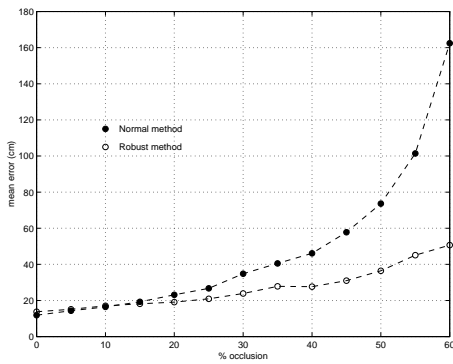


Figure 6. Mean error of localization for the standard and for the robust method.

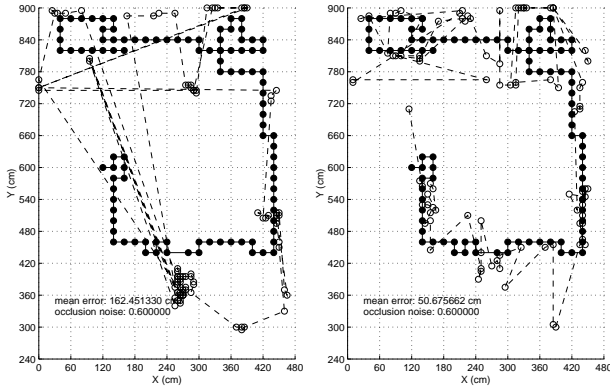


Figure 7. Localization on an imaginary path of 100 images at 60% occlusion. Left: standard method; right: robust method.

4. Experimental results

To perform the experiments we used a training set of 62 cylindrical panoramic images taken indoors in a laboratory with lots of occlusion and artificial lighting. The images were taken using a spherical mirror camera and warped to form cylindrical images. The images were taken at positions 60 cm apart. The experimental layout is depicted in Fig. 5, with squares denoting the positions where training images were taken. As a testing set we used 100 images taken at measured positions, depicted in Fig. 5 as full circles.

From the training set of images we constructed a 10 dimensional eigenspace. The empty circles in Fig. 5 denote the recovered path after projecting all 100 original test images. The spline used for projection was interpolated at 5 cm resolution. Since there is no significant occlusion (besides some change in the illumination of the windows area),

the standard and the robust method of coefficient retrieval perform almost equally well regarding precision. In fact, as it can be seen from the graph in Fig. 6, the mean error of the localization is between 11 cm and 13 cm for 0% occlusion. The performance of the robust estimator may vary slightly since the hypothesis generation includes a stochastic step.

The performance of both methods at higher levels of occlusion noise is compared in Fig. 6. We can see a significant improvement in precision as a result of applying the robust method. Even in situations of severe occlusion when more than half of the surrounding is invisible, the robust method retrieves positions that are reasonably close to the correct ones. This can be clearly seen in Fig. 7. On the left we can see that the standard method breaks under ambiguity of the data while the results of the robust estimator on the right show quite regular localization results with mean error under 60 cm.

5. Conclusion

In this paper we presented a method for robust view-based localization using panoramic images. As our experiments show, we can perform relatively accurate localization by using a pure view-based model of a pre-learned environment. By applying a robust framework to the recognition phase we can also achieve a significant improvement of performance when occlusions or noise are present in the input images. If we consider a scenario of a mobile robot in an office environment, the expected levels of noise seem acceptable for the algorithm.

We are currently exploring the problem of robustness in the learning phase and incremental on-line building of models of the environment.

References

- [1] H. Aihara, N. Iwasa, N. Yokoya, and H. Takemura. Memory-based self-localisation using omnidirectional images. In *14th ICPR*, pages 297–299, 1998.
- [2] H. Ishiguro and S. Tsuji. Image-based memory of environment. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pages 634–639, 1996.
- [3] M. Jogan and A. Leonardis. Panoramic eigenimages for spatial localisation. In *8th CAIP*, pages 558–567, 1999.
- [4] S. Judd and T. Collett. Multiple stored views and landmark guidance in ants. *Nature*, 392:710–714, 16 April 1998.
- [5] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *CVIU*, 78(1):99–118, 2000.
- [6] S. Nayar and T. Poggio, editors. *Early visual learning*. Oxford University Press, 1996.
- [7] S. K. Nayar, S. A. Nene, and H. Murase. Subspace methods for robot vision. *IEEE Trans. on RA*, 12(5):750–758, 1996.
- [8] T. Pajdla and V. Hlaváč. Zero phase representation of panoramic images for image based localization. In *8th CAIP*, pages 550–557, 1999.