Robust Localization Using Eigenspace of Spinning-Images *

Matjaž Jogan and Aleš Leonardis
Computer Vision Laboratory
Faculty of Computer and Information Science
University of Ljubljana
Tržaška c. 25 1001 Ljubljana, Slovenia
{matjaz.jogan,alesl}@fri.uni-lj.si

Abstract

Under in-plane rotations of a panoramic camera, the information content of a panoramic image is, in general, preserved. However, different representations that can be derived have important implications on further processing, e.g. for appearance-based localization. We discuss several approaches based on different representations that have been proposed and evaluate them from different points-of-view, in particular, we argue that most of them are not suitable for robust localization under partially occluded views. In this paper we propose a representation—eigenspace of spinning-images—which enables a straightforward application of the robust estimation of eigenimage coefficients which is directly related to the localization.

1. Introduction

Recent studies have shown good prospects for appearance-based localization using panoramic sensors [1, 6, 7, 8, 14]. Panoramic sensors capture a wide field-of-view and enable efficient characterization of a location by a single panoramic image[1]. During the training (learning) phase the panoramic images are collected in the environment. The redundant information captured in similar views is usually efficiently handeled by compressing the set of images with a PCA transform which leads to the so-called eigenspace representation [12, 13] of the environment map. Other basis functions such as Fourier basis [7] have also been used to approximate the learning set. The actual localization phase can be described as the search for the closest match between the current view and the views (or their interpolations) in the environment map.

The appearance-based localization strategy described herein is strongly related to the appearance-based object recognition [10, 12, 13]. However, in contrast to object recognition, where the target to be recognized usually occupies only a part of the image (on a cluttered background), in our case the complete image has to be recognized. When using panoramic images as representations of positions, we can expect that views taken at nearby positions and oriented in the same way tend to be strongly correlated. This allows us to build a compact representation that eliminates redundancy.

The information content of a panoramic image depends only on the location at which the image is captured and not on the orientation of the sensor. However, panoramic images taken at the same location may not be suitable for direct matching if they differ due to different in-plane orientations of the panoramic camera. The rotation variance must therefore be either encoded in the representation, or handled at the recognition stage. Usually panoramic images are being transformed to cylindrical panoramic images, so a change in the orientation of the sensor (in a plane perpendicular to the optical axis) results in a row-wise shift of the image (see Fig. 1).

In the literature one can find several approaches that try to deal with the problem of in-plane rotations of the panoramic sensor [1, 7, 14]. Driven by a desire to obtain a compact representation, the general tendency has been to represent each position with just one image. Having one image for each location in an arbitrary orientation in general prevents an efficient matching to estimate the localization. Thus, different representations have been proposed that achieve some sort of rotational-invarance, which translates in a shift-invariance for cylindrical panoramic images. One way to achieve the invariance is by applying a tranformation that produces the same output regardless of the shift in the input image. Such a transformation is

*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).

[1]In some applications even simplified one-dimensional surround views can be used [4, 5, 6].

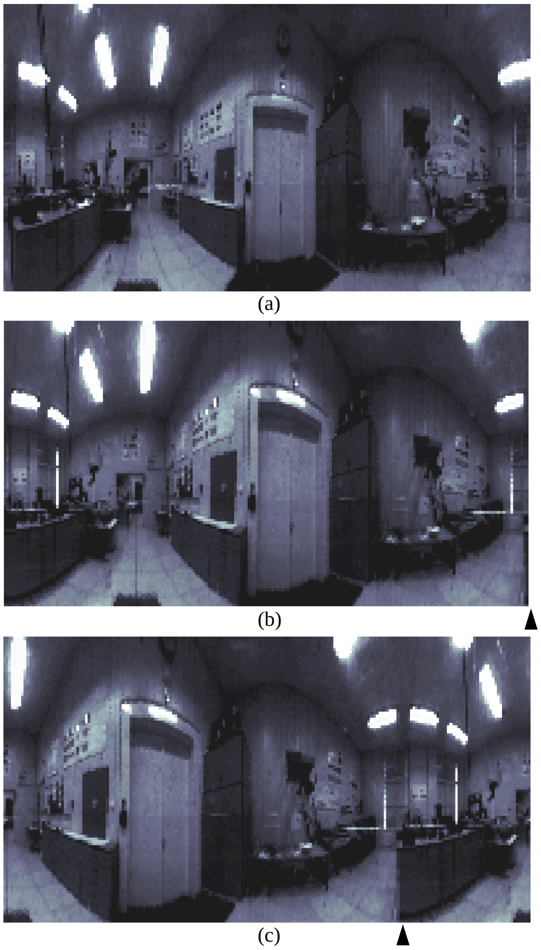

Figure 1. The cylindrical panoramic images (a) and (b) taken at viewpoints 60 cm apart. The image (c) was taken at the same position as (b), with sensor rotated by $90°$.

autocorrelation (which in the particular case of cylindrical panoramic images can also be applied row-wise [1]) or the FFT power spectrum. Another approach is to orient the images in a reference orientation. This can be achieved either by using some external sensors, e.g., a gyro-compass or exploiting the image content as it is the case of Zero-Phase-Representation (ZPR) [14].

We argue in this paper that while these approaches achieve shift-invariance, they fail at the localization stage when the input image locally deviates from the image stored in the environment map. These deviations can sometimes be a side effect of the design of the panoramic sensor (e.g., self-occlusion of the camera holder) or may be caused by occlusions due to objects moving in the environment. Since panoramic images capture a wide field-of-view it is almost impossible to avoid any disturbances during the localization (operation) phase, thus it is essential that a representation of the environment map enables efficient robust matching. Thus, we propose in this paper a representation—eigenspace of spinning-images—which enables a straightforward application of the robust estimation of eigenimage coefficients which is directly related to the localization.

The paper is organized as follows: In the next section we give a short overview of different representations that have been proposed, and point out their major features. In Section 3 we introduce our new representation which is based on the integration of all possible shifted (oriented) images in an eigenspace, i.e., *eigenspace of spinning-images*, and show some interesting properties of this representation. In Section 4 we describe the robust approach to estimation of coefficients of the eigenspace which is directly related to the localization problem. We conclude with a summary in Section 5.

2. Related work

In this section we give a short overview of the representations that have been proposed for the appearance-based localization using panoramic images and point out their major features. The representations are discussed with respect to their computational complexity, storage demands, and ability to be used for robust localization. Particularly, we emphasize robustness, since localization should be performed in dynamic environments, where noise and occlusions occur frequently.

A simple straightforward approach is to build a representation by taking one image at each location in an arbitrary orientation of the sensor. At the matching stage, the input image (captured in a different unknown orientation) must be sequentially shifted and matched, which is computationally very demanding if not prohibitive. However, robust localization *is* possible since the two images (the input image and the stored one) can be compared locally, which means that we can, at least in principle, successfully cope with occlusions [2].

Another approach is to perform autocorrelation, either by row and column or just by row direction, on both the training set of images and on the input images [1], which results in a shift-invariant representation. Recognition is then performed as direct matching between the input image and the stored ones. An example of a row-autocorrelated image is shown in Fig 2. However, the method has two drawbacks: first, the autocorrelation is not a one-to-one mapping and therefore different images can result in equivalent representations. Secondly, by autocorrelating an image local deviations are spread over the whole transformed image, therefore the method is non-robust with respect to partial occlusions in the input image.

Another approach is to use images that are oriented in

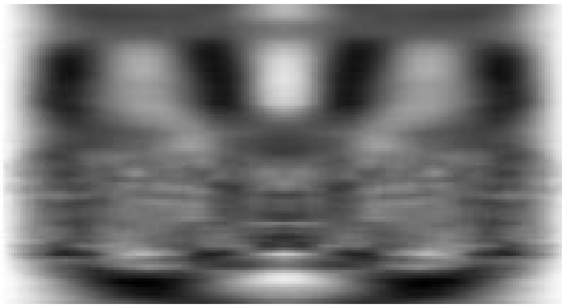

Figure 2. A row-autoccorrelated cylindrical panoramic image.

a reference orientation. For example, we can estimate the absolute orientation from other sensors, such as a gyrocompass, light polarization etc. These methods are sucessfull if such sensors are available and enough reliable. On the other hand, Pajdla and Hlaváč [14] suggest to estimate a reference orientation from images alone with the Zero Phase Representation (ZPR). ZPR, in contrast to autocorrelation, tends to preserve the original image content while at the same time achieving rotational invariance. It orients images by zeroing the phase of the first frequency of the Fourier transform of the image (see Fig. 3). The method is sensitive to variations in the scene, since it operates with only one frequency using a global transform. To alleviate the problem, one could apply ZPR along individual rows, to possibly improve the robustness in case that at least some of the rows are not occluded.

Figure 3. Left: original cylindrical panoramic images taken at positions 60 cm apart in random orientation. Right: Images from the left shifted as determined by the ZPR.

In our previous work [8] we have experimented with different representations in the eigenspace framework. We argued that representations that are reversible allow a higher localization performance. We have also shown that the different representations are comparable with respect to storage and computational complexity.

Realizing that the methods based on the ZPR and the autocorrelation are non-robust, we propose a representation which allows us to apply robust local methods for recognition.

3. Eigenspace of spinning-images

To enable efficient matching of input images which may be partially occluded, we propose in this paper a novel representation which incorporates the information about all possible orientations of the panoramic sensor (althought only one image per location needs to be taken). The principal idea is to acquire one image at each location in the learning stage and then shift the cylindrical panoramic image rowwise in order to simulate all possible rotations. All images generated in this way are then compressed by the PCA to form the final representation. Localization is then performed by projecting the momentary image directly onto the eigenspace (in the case of robust procedure we in fact have to solve an overconstrained set of equations [10, 11]), followed by searching for the nearest point on the spline.

The approach needs no preprocessing of individual images and the overall appearance is preserved in the representation. As the representation has all the desired properties, the question is how much the storage demands increase in this case. A graph showing the reconstruction accuracy (derived from the cumulative sum of eigenvalues) with respect to the percentage of eigenvectors for different number of stored images per location can be seen in Fig. 4. The representation with N images is built using one image per location, each of them oriented in a reference direction. The representations denoted by $N * K$ are made of images that are shifted sequentially by $360°/K$. As one can see, the compression ratio increases with the number of stored images per location. The graph in Fig. 5 shows the ratio of the storage cost and the processing cost for achieving the 20% and 40% reconstruction error with respect to the number of images stored per location. The processing cost is regarded as the number of shifts needed to compensate for the lack of orientations in the representation. When increasing the number of input images by 50x, the dimensionality of the eigenspace required to achieve same reconstruction error increases just by a factor of 2. This ratio even decreases with the increasing number of input images.

Features of the covariance matrix and eigenvectors

As we have shown, the set of shifted images compresses well and it can be efficiently represented at an acceptable cost in terms of storage requirements. It is interesting to analyze the covariance matrix produced by a set of shifted

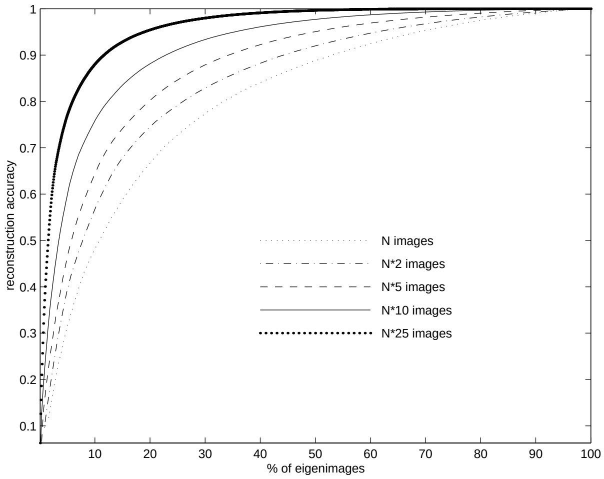

Figure 4. Reconstruction accuracy derived from the cumulative sum of eigenvalues.

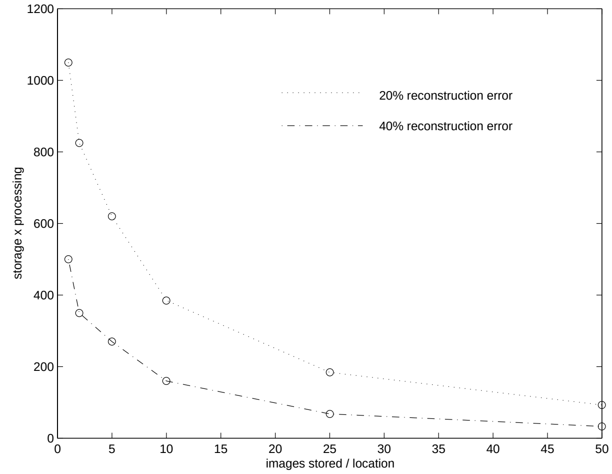

Figure 5. *Storage cost* $\times$ *processing cost* **for the** 20% **and** 40% **reconstruction error with respect to the number of images stored per location.**

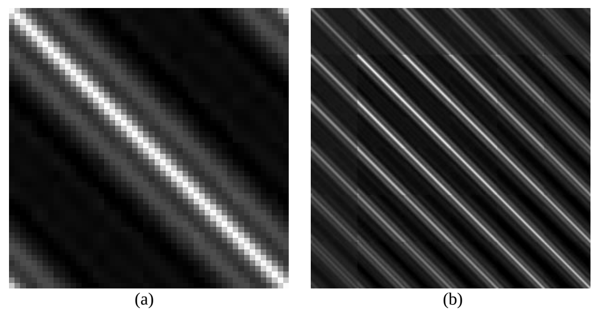

Figure 6. The covariance matrices: (a) $1D$ **signal and (b)** 3**-row image.**

images and the implications on the eigenvectors. It seems that these properties can help us further reduce the overall complexity.

Let us first analyze a 1-D case: we take a 1-D signal $\mathbf{x} = [x_0, x_1, \ldots, x_n]$ and then shift it sequentially in order to create a spinning-signal matrix

$$X = \begin{bmatrix} x_0 & x_1 & \ldots & x_n \\ x_n & x_0 & \ldots & x_{n-1} \\ x_{n-1} & x_n & \ldots & x_{n-2} \\ \ldots & & \ldots & \ldots \\ x_1 & x_2 & \ldots & x_0 \end{bmatrix} . \qquad (1)$$

Next we perform a PCA transform. To obtain the eigenvectors, we have to calculate the covariance matrix $Q = XX^T$. It can be shown, that, when the matrix X contains all the possible shifts of the signal, the covariance matrix forms a symmetric Toeplitz matrix:

$$Q = \begin{bmatrix} R_0 & R_1 & R_2 & \ldots & R_n \\ R_1 & R_0 & R_1 & \ldots & R_{n-1} \\ R_2 & R_1 & R_0 & \ldots & R_{n-2} \\ \ldots & & \ldots & \ldots & \\ R_n & R_{n-1} & R_{n-2} & \ldots & R_0 \end{bmatrix} . \qquad (2)$$

The eigenvectors of the Toeplitz matrices have some interesting properties [3]:

Theorem 1 Let Q have distinct eigenvalues. Then Q has $\lceil n/2 \rceil$ symmetric and $\lfloor n/2 \rfloor$ skew symmetric eigenvectors.

Theorem 2 If the eigenvalues of Q are distinct and arranged in a descending order, the corresponding eigenvectors will be alternately symmetric and skew symmetric. If n is odd, the first vector will be skew symmetric. If n is even, the first eigenvector can be either symmetric or skew symmetric.

Based on this two theorems, one can further decrease the storage requirements.

Now let us analyze a case of a spinning-image. A 2-D image consists of several rows that are individually shifted and the matrix of spinning-images looks like

$$X = \begin{bmatrix} x_0 & x_1 & \ldots & x_n & y_0 & y_1 & \ldots & y_n \\ x_n & x_0 & \ldots & x_{n-1} & y_n & y_0 & \ldots & y_{n-1} \\ \ldots & & \ldots & \ldots & \ldots & & \ldots & \ldots \\ x_1 & x_2 & \ldots & x_0 & y_1 & y_2 & \ldots & y_0 \end{bmatrix} .$$

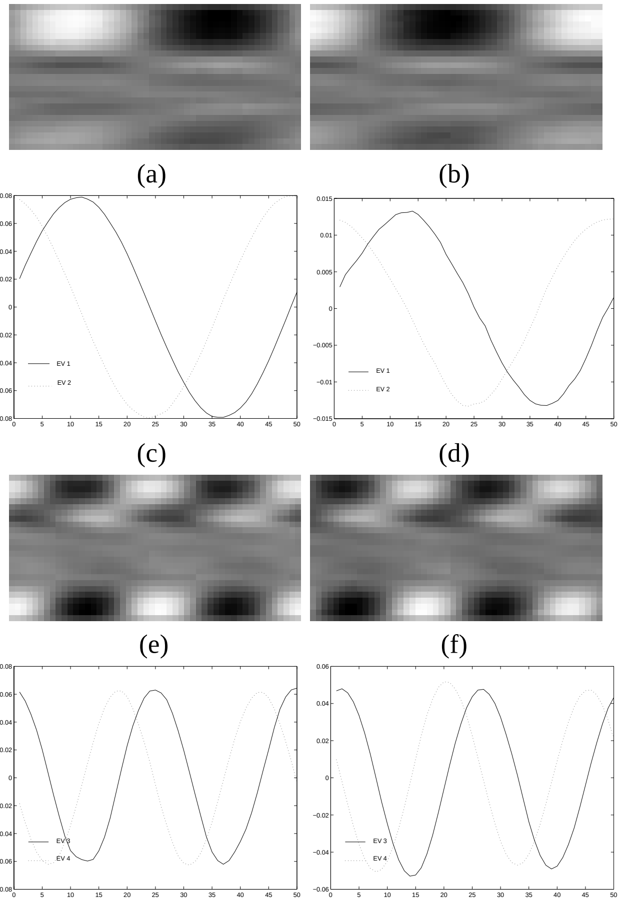

Figure 7. (a,b) The first two eigenvectors (eigenimages) and their second (c) and twentieth (d) rows. The third (e) and the fourth (f) eigenvector and their second (g) and twentieth (h) rows.

The covariance matrix of X in this case consists of blocks, which are Toeplitz matrices.

Two examples of covariance matrices are shown in Fig. 6, one for a 1-D signal (a) and one for a 3-row image (b).

The properties of the eigenvectors of the Toeplitz matrices cannot be directly applied to the 2D cylindrical panoramic images. However, when analyzing the properties of our experimental results, we found that the shape of the eigenvectors row-wise follows the Theorems 1 and 2 to a significant degree. As it can be seen from Fig. 7, the principal eigenvectors (depicted as eigenimages) take the shape of harmonic functions of different frequencies. The functions appear in pairs, as suggested by the Theorem 2.

The distribution of coefficients in the eigenspace (see

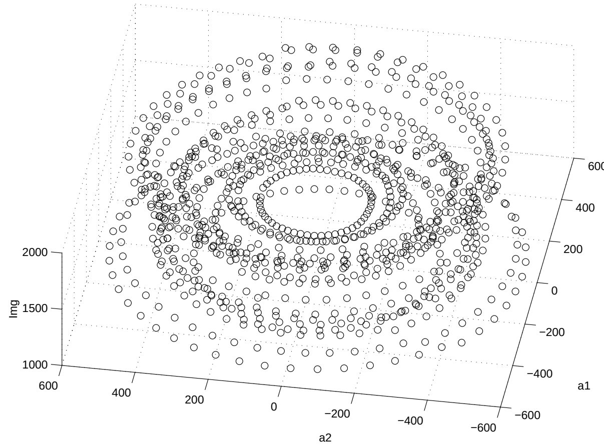

Figure 8. The distribution of coefficients in a $2-D$ space spanned by the first two eigenvectors.

Fig. 8) also raises questions about the possibility of an alternative computation technique, which will be explored in our future work.

Localization experiments

The results on the localization using the proposed method are presented in Fig. 9. The environment of the CMP lab was represented by a learning set of 62 cylindrical panoramic images, taken at positions on squares of 60×60 cm [9]. These positions are denoted as small squares in Fig. 9. The images were taken using a panoramic camera with a spherical mirror and warped to form cylindrical images. Then, all the images were sequentially shifted in steps of $7.2°$. The principal eigenvectors were calculated and selected as bases for the eigenspace consisting of the projections of all the images (62 acquired and 3038 obtained by shifting the original image). The points so obtained were interpolated on a 5×5 cm grid to form a spline. We tested the localization by using a 2, 10, and 20-dimensional eigenspace. As a testing set we used 100 images taken at measured positions, depicted in Fig. 9 as full circles[2]. There are no significant occlusions in the images of the testing set, besides some changes in the illumination in the windows area. The empty circles in Fig. 9 denote the recovered path after projecting all 100 original test images. As it can be seen from the figure, we can achieve a mean error of under 15 cm by using a 20-dimensional eigenspace. Next we explain how to perform the localization in case of partial occlusions in the input images.

[2]None of the images of the learning set was included in the testing set.

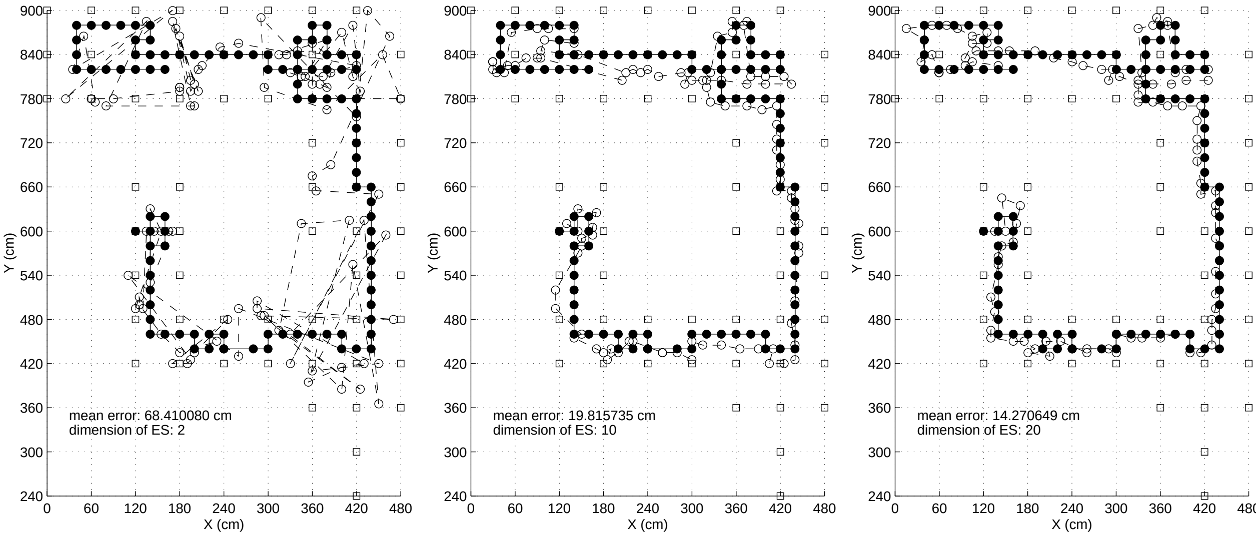

Figure 9. Localization test on 100 images using a 2, 10 and 20 dimensional eigenspace.

4. Robust recognition

Once the model (i.e., the eigenspace) is built, recognition of a view is performed by recovering the coefficient vector $\mathbf{q}$ which defines the linear combination of eigenvectors that characterize the input image. As every point in the eigenspace is associated with 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 p, $p \leq n$-dimensional[3] eigenspace [13]:

$$q_j(\mathbf{y}) = <\mathbf{y}, \mathbf{e}_j>; \;\; j = 1 \ldots p \;\; , \tag{3}$$

where $\mathbf{y}$ is the novel image vector and $\mathbf{e}_j$ are the eigenvectors. However, such calculation of parameters is non-robust and thus not successful in the case of noisy or occluded input 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 a robust approach [10, 11], that, instead of using the complete image vectors, generates and evaluates a set of hypotheses $\mathbf{r}$ as subsets of image points $\mathbf{r} = (r_1, r_2, \ldots, r_k)$. In fact, the coefficients can be recovered by solving a set of linear equations on $k = n$ points:

$$x_{r_i} = \sum_{j=1}^{n} q_j(\mathbf{x}) e_{jr_i} \;\; 1 \leq i \leq n \;\; . \tag{4}$$

The principle of such computation is illustrated in Fig. 10.

[3] n denotes the full dimensionality of the eigenspace.

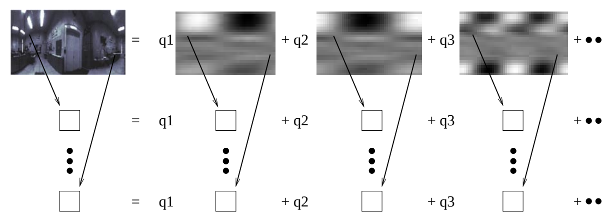

Figure 10. Calculating the coefficients from a system of linear equations.

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

$$E(\mathbf{r}) = \sum_{i=1}^{k} (x_{r_i} - \sum_{j=1}^{p} q_j(\mathbf{x}) e_{jr_i})^2 \;\; . \tag{5}$$

We solve the system on $k, k > p$ points, where k is significantly smaller than the total number of image points. We first randomly initiate the set of k points and then seek the solution which minimizes Eq.(5) in a least squares manner. Then we repeatedly reduce the set of points by eliminating those with the largest error. By doing so we achieve that only the points with a small error contribute to the computation of parameters. As we can see in Fig. 11, at the end most of the points in the occluded regions are excluded from the computation.

To increase the probability of avoiding points that are noise or belong to occlusion, several different hypotheses are generated. A hypothesis consists of a set of parame-

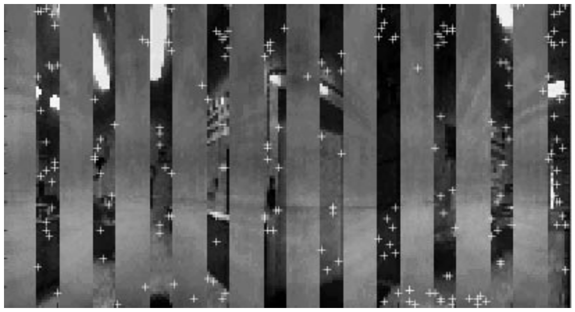

Figure 11. 60% occluded image. Crosses denote the points that contribute to the generation of a hypothesis.

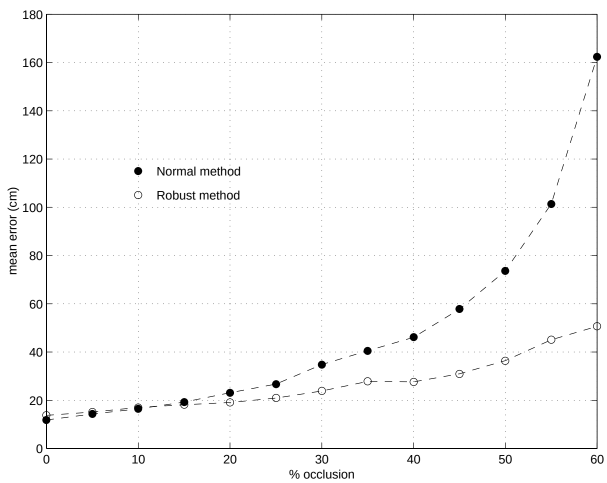

Figure 12. Mean error of localization for the standard and for the robust method.

ters, an error vector $\boldsymbol{\epsilon}$ 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 [10, 11]. The selection procedure selects the good hypotheses and rejects the superfluous ones by promoting those with a large number of encompassed points and a small overall deviation between the data and the reconstruction.

The performance of both the standard and the robust method at higher levels of occlusion noise is compared in Fig. 12. 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 image is occluded, the robust method retrieves positions that are reasonably close to the correct ones. This can be clearly seen

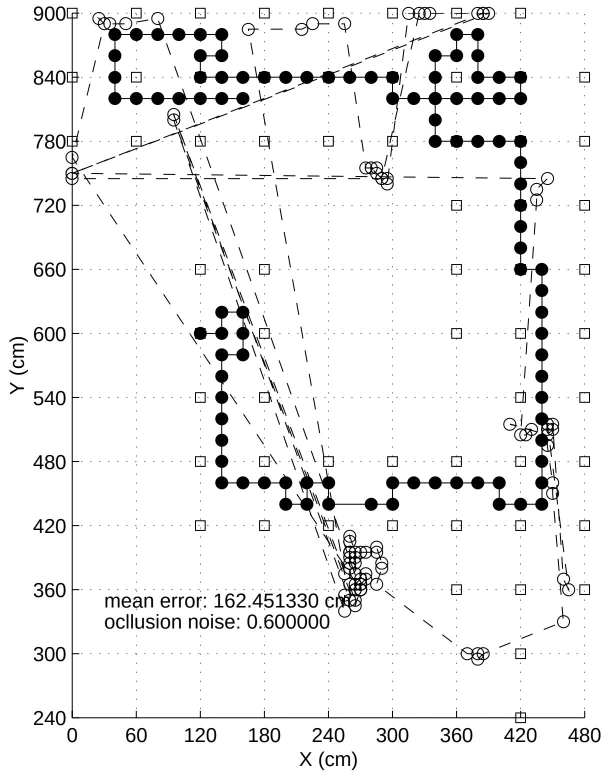

Figure 13. Localization on an imaginary path of 100 images at 60% occlusion with the standard method.

in Figs. 13 and 14. On the left we can see that the standard method breaks down 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. Conclusions

In this paper we have presented a representation—eigenspace of spinning-images—which leads to an efficient and reliable appearance-based localization using panoramic images. We have shown that the gains outweigh by far the moderate increase in the storage requirements with respect to the previous representations. The major strength of the representation is that it enables robust localization under partial occlusions which has been demonstrated by the experiments on a large data set.

Our work in progress is directed towards two goals: first to improve or modify the proposed representation based on the specific features revealed by the analysis of covariance matrix and eigenvectors of the covariance matrix obtained for a set of shifted images, and secondly to improve the robust procedure of estimating the coefficients of eigenimages

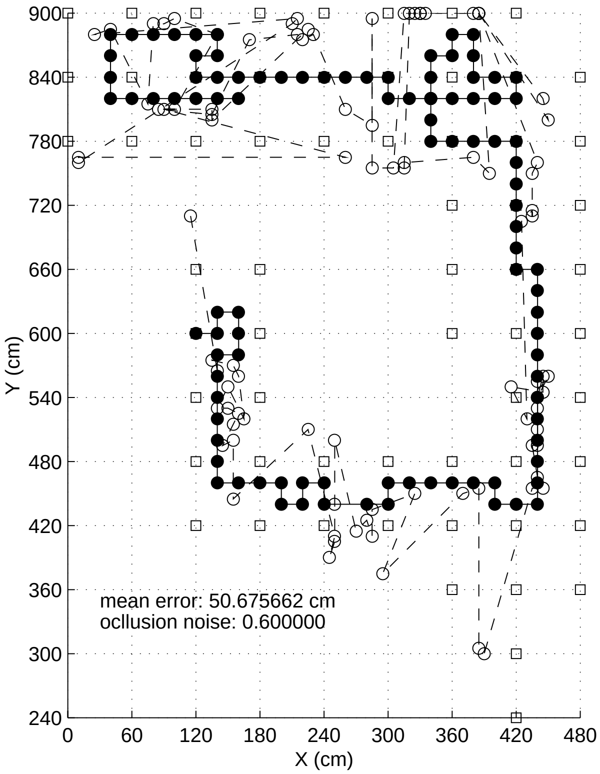

Figure 14. Localization on an imaginary path of 100 images at 60% occlusion with the robust method.

by exploiting local structure of panoramic images.

6. Acknowledgements

Matjaž Jogan acquired the images during his visit to the Center for Machine Perception, Czech Technical University, Prague. He would like to thank the staff of the CMP for their kind assistance. The visit was supported by the SI-CZ Intergovernmental S&T Cooperation Programme. The authors would also like to thank Horst Bischof and Tomáš Pajdla for fruitful discussion.

References

[1] H. Aihara, N. Iwasa, N. Yokoya, and H. Takemura. Memory-based self-localisation using omnidirectional images. In A. K. Jain, S. Venkatesh, and B. C. Lovell, editors, *14th International Conference on Pattern Recognition*, pages 297–299. IEEE Computer Society Press, August 1998.

[2] R. Brunelli and S. Messelodi. Robust estimation of correlation with applications to computer vision. *PR*, 28(6):833–841, June 1995.

[3] A. Cantoni and P. Butler. Properties of the eigenvectors of perysimmetric matrices with applications to communication theory. *IEEE Transactions on Communications*, COM-24(8):804–809, August 1976.

[4] J. Chahl and M. Srinivasan. Visual computation of egomotion using an image interpolation technique. *Biol.Cybern.*, 74(5):405–411, 1996.

[5] N. Franceschini, J. Pichon, and C. Blanes. From insect to robot vision. *Phil. Trans. R. Soc. Lond.*, (B 337):283–294, 1992.

[6] M. Franz, B. Schölkopf, H. Mallot, and H. Bülthoff. Where did I take that snapshot? Scene-based homing by image matching. *Biol. Cybern.*, 79(3):191–202, 1998.

[7] H. Ishiguro and S. Tsuji. Image-based memory of environment. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pages 634–639, 1996.

[8] M. Jogan and A. Leonardis. Panoramic eigenimages for spatial localisation. In F. Solina and A. Leonardis, editors, *8-th International Conference on Computer Analysis of Images and Patterns*, number 1689 in Lecture Notes in Computer Science, pages 558–567. Springer Verlag, September 1999.

[9] M. Jogan and T. Pajdla. Image-based map of the lab 130. Research Report CTU-CMP-1999-23, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, December 1999.

[10] A. Leonardis and H. Bischof. Dealing with occlusions in the eigenspace approach. In *Proc. of CVPR 96*, pages 453–458. IEEE Computer Society Press, 1996.

[11] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding - Special Issue on Robust Statistical Techniques in Image Understanding*, 2000. In press.

[12] H. Murase and S. Nayar. Visual learning and recognition of 3-d objects from appearance. *IJCV*, 14(1):5–24, January 1995.

[13] S. K. Nayar, S. A. Nene, and H. Murase. Subspace methods for robot vision. *IEEE Trans. on Robotics and Automation*, 12(5):750–758, October 1996.

[14] T. Pajdla and V. Hlaváč. Zero phase representation of panoramic images for image based localization. In F. Solina and A. Leonardis, editors, *8-th International Conference on Computer Analysis of Images and Patterns*, number 1689 in Lecture Notes in Computer Science, pages 550–557. Springer Verlag, September 1999.