Panoramic Eigenimages for Spatial Localisation *

Matjaž Jogan and Aleš Leonardis

Computer Vision Laboratory
Faculty of Computer and Information Science
University of Ljubljana
Tržaška 25, 1001 Ljubljana, Slovenia
e-mail: matjaz.jogan@mailcity.com

Abstract. Recent biological evidence suggests that position and orientation can be estimated from an adequately compressed set of environment snapshots and their relationships. In this paper we present a pure appearance-based localisation method using an eigenspace representation of panoramic images. We first review several types of rotational invariant representation of panoramic images in terms of their efficiency for an eigenspace-based localisation problem. Then, for each set of images an eigenspace from 25 location snapshots is built and analyzed. We evaluated simple localisation of images not included in the training set. The results show good prospects for the panoramic eigenspace approach.

1 Introduction

It is well known that a large number of animal species uses predominantly visual information to navigate in space. Most animals use visual information in combination with odometry, but in special cases, such as moving in the air or underwater, this is not possible. Several methods have therefore been implemented that use only vision to navigate. Nelson and Aloimonos [13] proposed that omnidirectional views could ease the task of estimating motion parameters and facilitate orientation. It seems reasonable to use omnidirectional views also for the localisation problem. Examples of using such an input can be found in the works of Yagi et al. [18] or Aihara et al. [1]. Simplified line-scan panoramic representations were proposed by Franz et al. [7], Francheschini et al. [6] and Chahl, Weber, Venkatesh and Shrinivasan [19, 3].

In pure appearance-based methods for navigation, which use simplified representations of views such as line-scan intensity rings [7, 19, 3], appearance cues are stored in memory to represent the explored space. Localisation is then performed by matching current views with those stored in the memory.

Recently, more evidence has been gathered which shows that, at some level, biological systems also perform appearance matching. Judd and Colett [9] reported that ants use multiple stored snapshots to learn a path from start to

* This work was supported by the Ministry of Science and Technology of Republic of Slovenia (Project J2-0414).

goal. Dill, Wolf, and Heisenberg [4] also reported that bees try to match retinotopically the incoming visual pattern with previously stored images. Appearance cues are found to serve either for matching or to be organized in higher cognitive schemes such as maps, as it is presumed for the hippo-campus brain area of mammals (e.g. Epstain [5]).

If we take instead of an intensity ring an iconic representation of the world, as in the case of snapshots taken with a wide visual field, it would be cumbersome to densely scan the environment and then perform matching operations to search for the most similar snapshot in the memory. In fact, because of the computational complexity and the amount of memory needed, this is not feasible even for a biological vision system. It is therefore obvious that the data must be compressed. An efficient method is the *Singular Value Decomposition (SVD)*, which can be used to find a low-dimensional representation of a set of images in terms of linear combinations (points) in eigenspace, spanned by orthogonal eigenvectors (eigenimages). If the data set distribution can be encompassed with a small number of eigenvectors, we can achieve significant dimensionality reduction. We can assume that two equally oriented images taken at close positions appear very similar, therefore we expect that (a) we can achieve a significant compression and (b) points in the eigenspace, representing neighboring positions, will also be close to each other. The SVD and similar methods have been widely used in appearance-based recognition problems [14, 17, 12, 2, 11].

The most straightforward application of eigenspaces therefore requires a representation of the visual input that would be easily aligned for matching. One way is to estimate the orientation from other sources such as light polarization, gyrocompass etc. We can also find a representation that would be rotation-invariant, such as row-correlation [1], but we loose the orientation specific information. An approach called *Zero Phase Representation (ZPR)* was recently proposed by Pajdla [15] which can be used to find a transformation that projects differently oriented but otherwise identical panoramic images into one representative image.

This paper is organized as follows. In section 2 we first review various rotationally invariant representations and evaluate them on our image set. In section 3 we describe the procedure of building eigenspaces with the SVD method and, finally, we present the results of localisation experiments in section 4. We conclude with a summary and outline future work.

2 Shift-invariant representations of panoramic images

2.1 Evaluation of panoramic images

Fig. 1 shows the map of the CMP Laboratory[2] with some examples of panoramic images taken at positions 1, 5, 12, and 15. From the original images, taken at random orientations, four sets of rotational independent images were generated.

[2] All panoramic images used in this paper were kindly provided by T.Pajdla and J.Černík from the Czech Technical University at Prague.

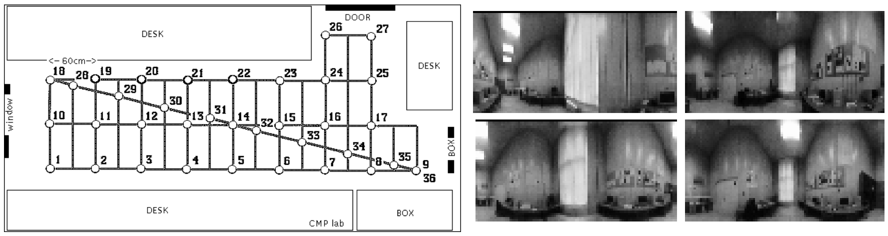

Fig. 1. Left: Map of the CMP lab with positions where the snapshots were taken. Right: Snapshots taken at positions 1, 5, 12, and 15. The center of each snapshot points towards West.

As our intention is to compress the image set by estimating the most significant eigenvectors and then present the training set as points in the eigenspace, we need a criterion to evaluate the image sets. A suitable criterion is the distance in the eigenspace. It has been shown [12, 14] that by estimating the *correlation* $i_p^T i_q$ of two images the distance between their projections in the eigenspace can be evaluated.

2.2 Manually aligned images

Manually aligned images represent the case in which the orientation of the sensor at the time of taking the snapshot is known and therefore the images can be shifted so that they are all oriented in the same direction. Since the orientation is the same, it is expected that images taken at small distances apart are strongly correlated, i.e., their correlation coefficient being higher than for images taken far apart. In Fig. 2 we can see how the correlation with the image obtained at position 30 varies with position for aligned images.

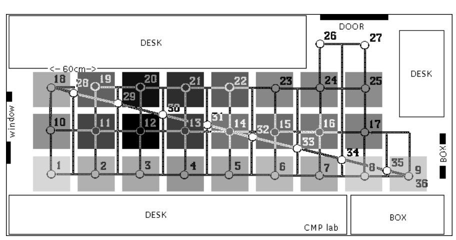

Fig. 2. Correlation with the snapshot taken at position 30 for manually aligned images. Darker areas indicate a higher level of correlation.

2.3 Autocorrelated images

Another way of achieving rotational invariance is by autocorrelating the images. We have two options. The first option is to perform autocorrelation by row and column directions, which results in not only rotational but also vertical invariance. It can be seen in Fig. 4(left) that the proximity of images based on the correlation of autocorrelated images seems weaker, compared with the performance of aligned sets. The second option is to correlate images just by row direction, which appeared in a similar context in [1]. It can be seen from Fig. 4(right) that the correlation values of row autocorrelated images can be compared to that of the fully autocorrelated ones.

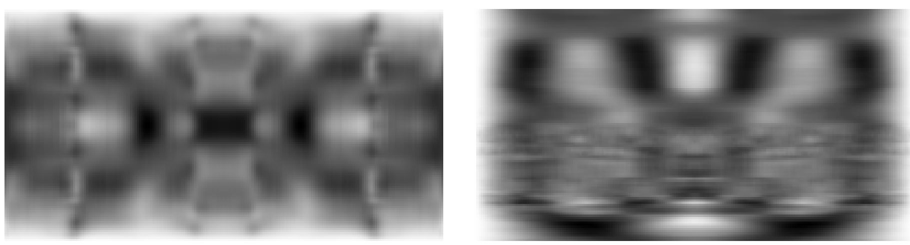

Fig. 3. Left: fully autocorrelated transform of the snapshot taken at position 1; Right: row autocorrelated transform of the same snapshot.

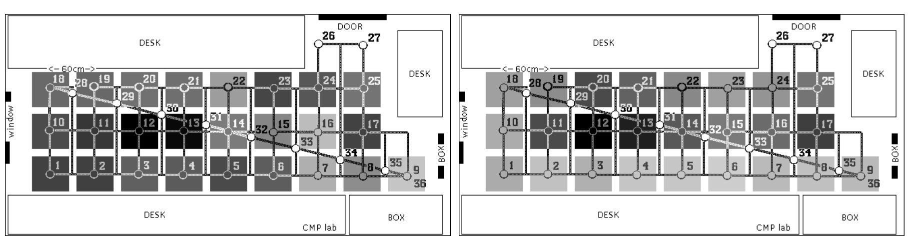

Fig. 4. Left: Correlation with snapshot taken at position 30 for fully autocorrelated images. Right: correlation with same snapshot for row autocorrelated images.

2.4 ZPR images

Zero Phase Representation (ZPR) images are obtained from original randomly oriented images by shifting the image so that its first harmonic in the *Discrete Fourier Transform* has a phase equal to zero [15]. This transformation produces one representative image for an equivalence class which, in our case, is the set of panoramic images taken under stable conditions at the same place but with possibly different orientations. If a panoramic image is represented as a 2-dimensional discrete function $I(row, col)$, then its ZPR representation $I_{ZPR}(row, col) = I(row, col - \phi)$ can be determined as

$$I_{ZPR}(row, col) = \mathcal{F}^{-1}\{\mathcal{F}\{I(row, col)\}e^{-j\Phi[F(0,1)]l}\} \ , \tag{1}$$

where $F(k, l) = \mathcal{F}\{I(row, col)\}$ is a Fourier transform of I and $\Phi[F(0, 1)]$ denotes the shift ϕ of the image necessary to obtain an image whose first harmonic has a phase equal to zero [15].

Problems may arise due to the uneven resolution of the cylindrical panoramic image caused by the transformation process from the original panoramic image and because of possible self-occlusion of image content on the top and bottom of the images. Thus one needs to robustify the method by weighting the images [16], as it can be seen in Fig. 6(left). In Fig. 5 we can see how the ZPR performs on panoramic images taken at positions 6 and 7. It is expected that representative images of equivalence classes taken not far apart would be strongly correlated and thus close in the parametric eigenspace. Our tests show that this expectation is correct since our set of 35 images, when transformed by ZPR, results in a set of very similarly oriented images. In Fig. 6(right) one can see that the correlation distribution can be compared with that of the manually oriented set.

Fig. 5. Left: original snapshots taken at positions 6 and 7 in random orientation. Right: ZPR transforms of the images on the left.

3 Panoramic eigenspaces

3.1 Construction

Eigenspaces were constructed from 25 panoramic snapshots taken from the positions on the rectangular grid, labeled 1-25, resprectively. For testing, we used the remaining images.

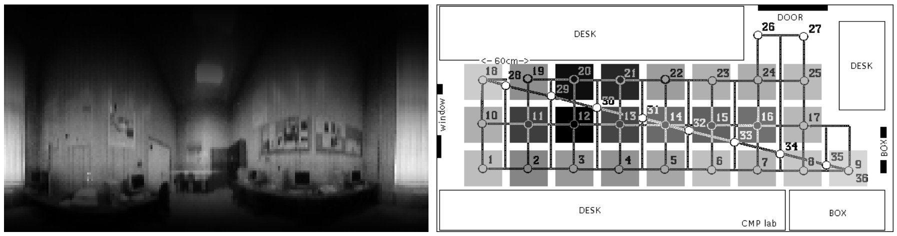

Fig. 6. Left: Weighted image taken at position 30. Right: Correlation with the snapshot taken at position 30 for weighted ZPR images. Darker areas indicate a higher level of correlation.

To reduce the computational complexity, eigenspaces were built with an algorithm that estimates the SVD of the smaller covariance matrix, that is, if A_{mn} is the matrix of n image vectors (normalized so that mean is in the origin), then the SVD of Q_{nn} is calculated:

$$Q_{nn} = A_{nm}^T A_{mn} = U_{nn} V_{nn} U_{nn}^T \quad . \tag{2}$$

The eigenvectors obtained with further processing of the matrix U_{nn} are then sorted according to their corresponding eigenvalues from the diagonal of V_{nn}.

3.2 Comparison of eigenspaces

We can define the *energy* of an eigenvector as proportional to the magnitude of its eigenvalue. Then we can estimate the *energy* of an eigenspace as the cumulative sum of the energies of the eigenvectors included. If we then take an eigenspace of dimension p, $p < n$, we can relate the compression rate to the energy of the eigenspace.

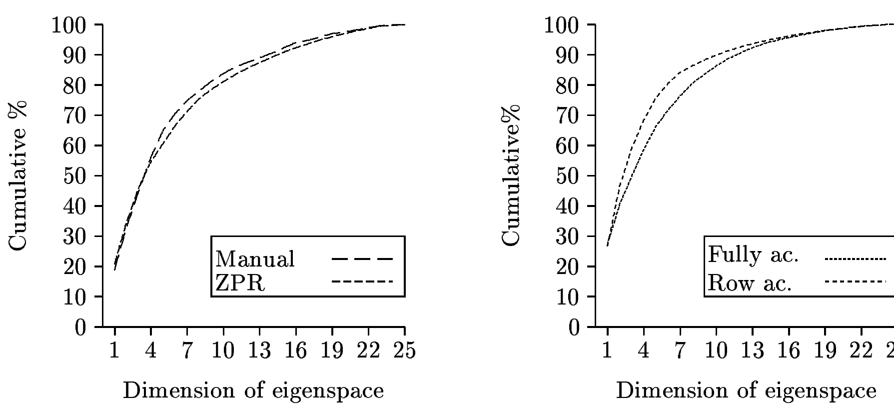

Fig. 7. Cumulative energy distribution for: manually aligned and ZPR of weighted image set (left); autocorrelated sets (right).

From the cumulative energy plots in Fig. 7, we can see the difference of energy distribution for eigenspaces based on manual and ZPR sets. ZPR performs slightly worse in terms of compactness of representation, which is mostly due to a greater dissimilarity between images taken from neighbor positions since orientation is not exactly uniform through the set.

Next, we compare the plots Fig. 7(left) with Fig. 7(right) and observe that we can for the correlated sets more energy can be captured in the first few eigenvectors. This would be favorable if our goal was to achieve a high compression rate, however when discrimination between the images is the primary objective, this is not necessarily the case. As it was previously stated, in the case of the autocorrelated image set similarity between images is not in strong relationship with proximity of the locations where the snapshots have been taken.

4 Experimental results of localisation

After the eigenspaces were built and analyzed we tested their performance in the localisation task. The test images were projected onto the eigenspace. As the correlation in Hilbert space equals the Euclidean distance in eigenspace [14, 12], the nearest neighbor can be used to estimate the most similar snapshot. We defined four criteria to measure the success rate:

I The **first criterion** tells whether the image from the training set representing the position nearest to that of the test image was successfully recognized as the nearest one.

II The **second criterion** tells whether the first *and* the second nearest positions were successfully recognized.

III The **third criterion** tells whether the *three* nearest positions were successfully recognized.

IV The **fourth criterion** is the weakest of all and tells whether *at least one* of the *four* nearest positions was recognized as the nearest one.

The first three conditions happen to be very strict since we use for the testing the images that were not included in the training set and are therefore taken at different (intermediate) positions. As we can see from Fig. 1(left), these positions, except for those numbered 26 and 27, lie on the diagonal line and comprise positions numbered from 28 to 35. In our case, there are some positions that lie near the middle of the quadrant and the difference of lengths to the nearest neighbors is small. It is obvious that if the first criterion fails, so do the second and the third. The fourth criterion however, checks only if one of the four positions surrounding the test position appears as the nearest.

In Fig. 8 we compare the performance on the manually aligned and the ZPR set of weighted images. The results for the manually aligned set seem good enough for approximate localisation. If we operate with a six-dimensional eigenspace, we correctly determine the three nearest positions (ordered) in 50% of the cases and the nearest position in 100% of the cases. An exact position estimation would of course require further processing and will not be discussed

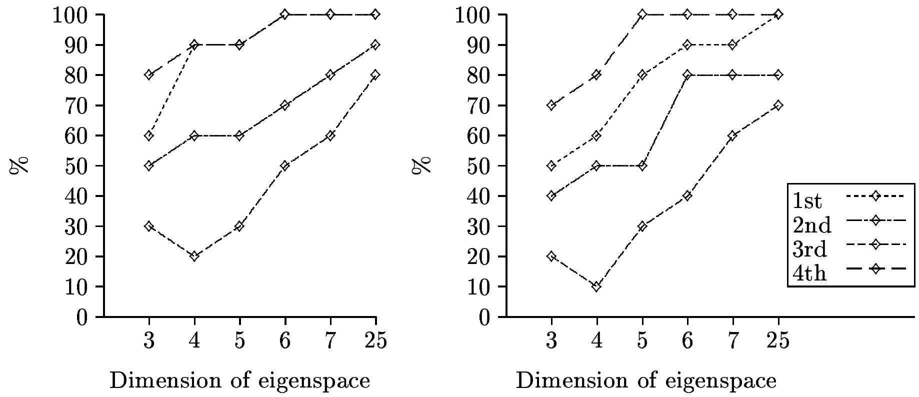

Fig. 8. Success rate for manually aligned (left) and ZPR of weighted image set (right).

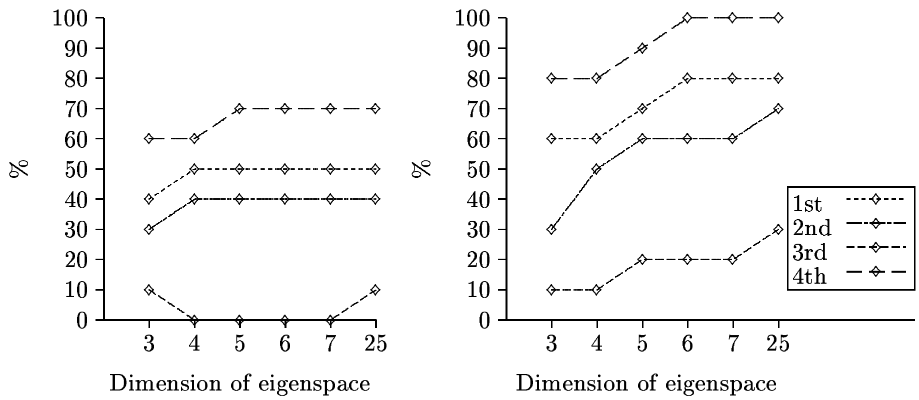

Fig. 9. Success rate for fully autocorrelated (left) and row autocorrelated image set (right).

in this paper. In Fig. 8(right) we can see that basically the same performance can be achieved by using ZPR transformed weighted images. As we can observe, eigenspaces of dimension less than 10 perform well enough and any further increase in the dimension leads to higher cost/performance rate.

In Fig. 9 we can see the same graphs for the fully autocorrelated and row autocorrelated images. As we can observe, the performance is not nearly as good as in the previous cases and it does not improve even with full dimensionality of the eigenspace.

In Fig. 10(left) we can see the result of position estimation for the snapshot number 30 from the manually aligned set of images. In Fig. 10(right) we can see how the method behaves while localising a snapshot taken at position 30 for the ZPR transformed weighted images. In this case, the result is as good as in the case of the manually aligned set.

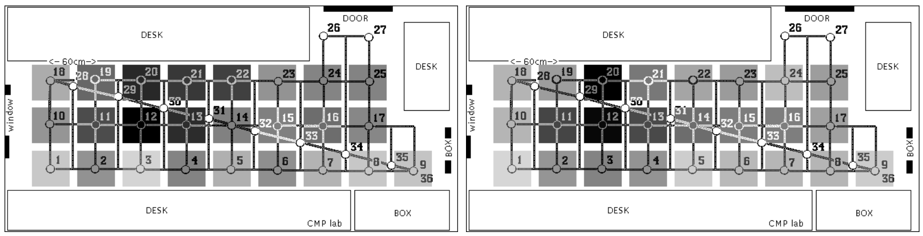

Fig. 10. Left: Localisation of snapshot 30 for the manually aligned set of images. Right: Localisation of snapshot 30 for the ZPR transformed weighted set of images. Darker shades denote smaller distance in the eigenspace.

5 Summary

In this paper we discussed the problem of appearance-based localisation. The eigenspace approach proves itself as a very viable one. Our tests show that the nearest positions can be estimated with significant certainty if we rely on similarity of equally oriented neighboring images. To achieve similar orientation of images taken with random orientation of the sensor, we tested the Zero Phase Representation which performed well enough for our limited image set without occlusions. It is however a matter of future research to analyze how it performs on larger sets and under unpredictable circumstances. We also tested the autocorrelated rotational invariant representations which do not preserve overall appearance. Our results show that the performance of such representations is inferior, mostly due to the fact that the relationship between the image correlation and distance of positions where the images are taken is not so explicit, as in the case oriented images. Note that, as opposed to [1], tests were performed only on the snapshots that were not in the training set, i.e., they were taken at slightly different positions.

In future work we plan to test additional sets of images, some with occlusions and under different illumination. On a wider set of images, multiple eigenspaces will be built and analyzed. On more densely sampled images, points in eigenspace will be interpolated to form a hypersurface and its characteristics will be analyzed. Strategies for calculating the exact position from localisation scores will be developed.

References

1. Aihara H., Iwasa N., Yokoya N., Takemura H.: Memory-based self-localisation using omnidirectional images. Anil K. Jain, Svetha Venkatesh, Brian C. Lovell, editors, 14th International Conference on Pattern Recognition, Brisbane, Australia, volume I. IEEE Computer Society Press (1998) 297-299
2. Black M. J., Jepson A. D.: EigenTracking: Robust matching and tracking of articulated objects using a view-based representation. International Journal of Computer Vision 26(1) (1998) 63-84

3. Chahl, J.S., Srinivasan M.V.: Visual computation of egomotion using an image interpolation technique. Biol.Cybern. Vol.74, No.5 (1996) 405-411
4. Dill M., Wolf R., Heisenberg M.: Visual pattern recognition in *Drosophila* involves retinotopic matching. Nature Vol 365 (1993) 751-753
5. Epstain, R., Kanwisher N.: A cortical representation of the local visual environment. Nature Vol.392 (1998) 598-601
6. Francheschini N., Pichon J.M. and Blanes C.: From insect to robot vision. Phil. Trans. R. Soc. Lond. B 337 (1992) 283-294
7. Franz M.O., Schölkopf B., Mallot H.A., Bülthoff H.H.: Where did I take that snapshot? Scene-based homing by image matching. Biol. Cybern. Vol.79, No.3 (1998) 191-202
8. Guerrero, J.J., Sagues C.: Tracking features with camera maneuvering for vision-based navigation. Journal of robotic systems Vol.15, No.4 (1998) 191-206
9. Judd S.P.D., Collett T.S.: Multiple stored views and landmark guidance in ants. Nature Vol.392 (1998) 710-714
10. Kurata J., Grattan K.T.V., Uchiyama H.: Navigation system for a mobile robot with a visual sensor using a fish-eyelens. Review of Scientific Instruments Vol.69 No.2, 1 (1998) 585-590
11. Leonardis A. and Bischof H.: Dealing with occlusions in the eigenspace approach. Proc. of CVPR'96, IEEE Computer Society Press (1996) 453-458
12. Nayar S. K., Nene S. A., Murase H.: Subspace methods for robot vision. IEEE Trans. on Robotics and Automation vol.12, No.5 (1996) 750-758
13. Nelson R.C. and Aloimonos J.: Finding motion parameters from spherical motion fields (or advantages of having eyes in the back of your head). Biol. cybernetics Vol.58 (1988) 261-273
14. Oja E.: Subspace methods of pattern recognition. Research Studies Press, Hertfordshire (1983)
15. Pajdla T.: Robot localization using panoramic images. Norbert Bräendle, editor, Proc. of the Computer Vision Winter Workshop, Rastenfeld, Austria. (1999) 1-12
16. Pajdla T.: Personal correspondence (1999)
17. Turk M. and Pentland A.: Face recognition using eigenfaces. Proc. Computer Vision and Pattern Recognition, CVPR'91 (1991) 586-591
18. Yagi Y., Nishizawa Y. and Yachida M.: Map-based navigation for a mobile robot with omnidirectional image sensor COPIS. IEEE Trans. on Robotics and Automation Vol.11, No.5 (1995) 634-648
19. Weber K., Srinivasan M.V.: An insect-based approach to robotic homing. ICPR'98 (1998) 297-299

This article was processed using the LaTeX macro package with LLNCS style