Continuity properties of the appearance manifold for mobile robot position estimation

J.L. Crowley*, F. Pourraz

Project PRIMA-IMAG, Lab Gravir-Imag, INRIA Rhone-Alps, 655 Avenue de l'Europe, 38330 Montbonnot, France

Abstract

This paper describes the results of experiments using an appearance manifold of a scene to estimate the position and orientation of a mobile robot. These experiments use principal components of images of the environment as an orthogonal basis for representing appearance from different positions and orientations. A regularly sampled grid of images is projected into a set of N principal component images to produce a 3D manifold within an N dimensional appearance space.

The orthogonal basis serves as an index for generating hypotheses of position and orientation from a newly acquired image. The basis provides a similarity metric for matching images. Distance from the projection of an observed image to the manifold gives an estimate of the confidence of a position estimate. Reconstructing the scene from the manifold allows obstacle detection.

In these experiments, we examine the effects of variations in image resolution, number of basis images, and density of samples used for constructing the manifold. We investigate the use of interpolation over the appearance manifold in order to improve the precision of the estimated position and orientation. We demonstrate position estimation with 1D, 2D and 3D manifolds. Our results indicate that such an approach can provide a fast and reliable method for visual navigation. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Continuity properties; Mobile robot; Visual navigation

1. Introduction

The appearance manifold of an object or a scene is the continuous set of possible images of the object or scene ordered by systematic variation of image formation parameters. Parameters which may be used to define an appearance manifold include the relative position and orientation of the camera, the camera intrinsic parameters, and the illumination function (including color, intensity and angle). While an appearance manifold can, in theory, be represented in the space of possible images, practical implementation requires projecting onto a basis with a lower number of dimensions. Such a basis is provided by a subset of the principal components of the image set [1,2]. Murase and Nayar [3] demonstrated that changes in viewpoint around ordinary objects could be expressed as a closed contour (or 1D manifold) in a space of principal component images. Black and Jepson [4] showed that changes of configuration of hands during hand gestures could be represented and matched using contours in such a space. McKenna and Gong [5] demonstrated use of a 2D manifold of face images

for head pose estimation. Many other investigators have demonstrated vision systems in which appearance is represented by contours, surfaces or manifolds in a feature space defined by principal components of images. In the experiments in this paper, we investigate whether such an approach can be used for visual navigation on a mobile robot within two spatial and one orientation dimension.

Wallner has recently demonstrated position estimation and obstacle detection using an appearance manifold constructed from a scanning laser range finder [6,7]. In this approach, a set of range scans is matched to build a composite model of the environment. This model is then used to construct synthetic range scans over a dense grid representing two spatial and one orientation dimension. This set of range scans is analyzed using principal components analysis to define an orthogonal basis. The grid of scans is then projected into this basis to provide a sampled 3D manifold representing possible laser scans. Position and orientation for a newly acquired scan can be obtained by using the orthogonal basis space as an index into the manifold.

In this paper we investigate whether a similar approach can be applied to position estimation using images from a single camera mounted on a mobile robot. In these experiments we assume a constant illumination direction. Energy

---

* Corresponding author. Tel.: +33-4-76615396; fax: +33-4-76615210.
  E-mail address: james.crowley@inrialpes.fr (J.L. Crowley).

normalization is used to achieve robustness to variation in illumination intensity. Movement of the camera is constrained to a 2D plane, and thus described by a manifold over three variables representing position and orientation. Images taken at arbitrary positions in the scene are projected into the appearance space by inner product with the basis images. The nearest point on the manifold is computed by interpolation and used to determine the position and orientation from which the image was acquired. The distance between the projection of the image and its nearest point in the manifold provides an estimate of a confidence factor.

In the following sections we describe the construction of the pose manifold, and its use to estimate pose. We explore the effects of variations in image resolution, the number of orthogonal bases and the density of samples over the appearance manifold. We present experimental results for position estimation of a mobile robot moving along the 1D path as well as a 2D grid in a laboratory and in a hallway environment. These experiments demonstrate that the number of basis dimensions of the appearance space, the density of sample images in the pose manifold and the required image resolution depend on the continuity of the pose manifold. Continuity of the pose manifold depends on the complexity of the visual environment which is to be represented.

2. An orthogonal basis for the appearance manifold

The appearance of a scene is the set of all possible images observed from different viewing positions and angles. A manifold of images may be obtained by systematically observing the scene at regularly spaced view positions and orientations. Such regularly spaced images may be obtained by physically positioning a camera using a robot, or by producing synthetic images using a dense correspondence and reprojection algorithm, such as described by Quan [8].

Projecting the images into an orthogonal basis provides an important reduction in memory required for storage while also providing an efficient method for indexing and matching images [2]. An orthogonal basis for expressing images of a scene may be determined by principal components analysis. This technique uses the eigenvectors of the covariance matrix of an image set to build an orthogonal basis whose characteristic is to gather the largest variance on its first eigenvectors. Consequently, the first few dimensions of this space are sufficient to capture the significant variations in scene appearance. One of the properties of this space is that the distance between two points (e.g. projected images) is the best approximation, in the least-square sense, to the correlation between the images. In other words, distance within the space is proportional to the inner product of images.

Principal components analysis is a well-established technique to determine an orthogonal basis for a set of images. However, its use with very large sets of images imposes

some subtle computational problems. In order to consider these issues we begin by briefly reviewing the computations involved. Let us suppose that we have N sample images, $\{S_n\}$ composed of M pixels per image. Consider the images as forming a single long 1D vector $S(m)$. An average image, $\mu(m)$, is computed from this set and subtracted from each image to provide a set of zero mean images, $\{W(m)\}$

$$\mu m(m) = \backslash f(1, N) \backslash \imath \backslash su(n-0, N-1, \cdot) S_n(m)$$

$$W_n(m) = S_n(m) - \mu(m)$$

A $M \times M$ covariance matrix, A, may be computed from the zero mean images. Each coefficient a_{jk} represents the expected (or average) product of the jth and kth pixel.

$$a_{jk} = \backslash f(1, N) \backslash \imath \backslash su(n-0, N-1, \cdot) W_n(j) W_n(k) \qquad (1)$$

The principal components of the image set are obtained by solving the eigen-structure decomposition problem using an algorithm such as Householders method or SVD [9]. These algorithms will produce a rotation matrix R and diagonal matrix Λ_a such that

$$AR = R\Lambda_a \qquad (2)$$

The M columns of the matrix R are orthogonal direction vectors which rotate the covariance matrix A into a diagonal matrix Λ_a. The diagonal terms of Λ_a are the variance of the image set in each of these directions. Most algorithms for principal components analysis return R and Λ_a such that the diagonal elements are ordered from largest to smallest. The rank of the matrix A is limited by the number of pixels, M, and the number of sample images N. If $N < M$ then the last N–M diagonal terms of Λ_a will be very close to zero (essentially small random values due to round-off noise).

In most applications, the image size, M, is in the range 2^{12}–2^{18} pixels, while the number of images, N, is on the scale of 2^6–2^{12} images. The cost and difficulty of computing a M by M covariance matrix are avoided by replacing A with an N by N matrix, B, computed as the covariance of the image set. The coefficients b_{jk} are the inner products of the jth and kth images

$$b_{jk} = \backslash f(1, M) \backslash \imath \backslash su(m=0, M-1, \cdot) W_j(m) W_k(m) \qquad (3)$$

Solving the eigen-structure problem for B gives N by N rotation matrix Q and diagonal matrix Λ_b such that

$$BQ = Q\Lambda_b \qquad (4)$$

Because A and B represent the same image set, they have the same principal components. Multiplying Eq. (4) by W_n^T, as suggested by Ref. [1], gives

$$W_n^T BQ = W_n^T Q\Lambda_b$$
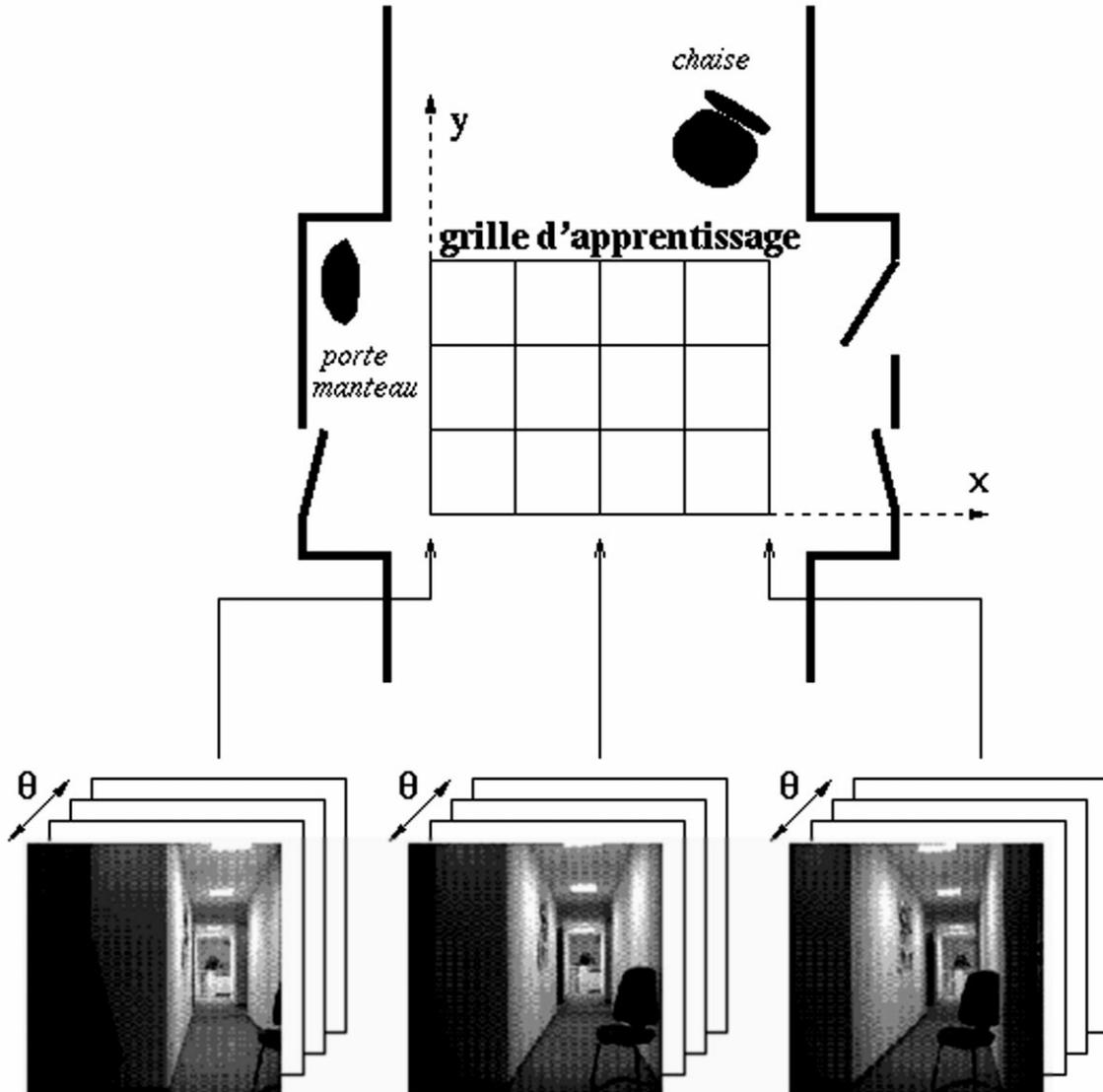
We note that

$$W_n^T B = AW_n^T$$

Fig. 1. In data set 4, images were obtained in a hallway at a 4 × 5 grid of positions spaced at 20 cm. For each position, 19 images were obtained with angular step of 10°.

So that

$$AW_n^{\mathrm{T}}Q = W_n^{\mathrm{T}}Q\Lambda_b = R\Lambda_a$$

Thus, the first N columns of R, can be obtained as a weighted sum of $W_n(m)$ using the columns of Q as weights

$$R = W_n^{\mathrm{T}}Q$$

If we define the kth column of R as $R_k(m)$, then

$$R_k(m) = \backslash f(1,N)\backslash i\backslash s\ \ \ (i\ \ = 0,\ N-1,\cdot)q_{jk}W_j(m) \tag{5}$$

These N columns provide N orthogonal bases, $R_k(m)$ for the image set $W_n(m)$.

Using either Eqs. (1) and (2) or Eqs. (3) and (5) we choose the first K columns from R as an orthogonal basis. We normalize the energy of each vector so that the basis is orthonormal. After such normalization, any image, $S(m)$, from the set S may be projected onto this basis with an inner product giving a vector of coefficients

$$\alpha_k = <(S-\mu), R_k> = \backslash f(1,M)\backslash i\backslash su(m$$
$$= 0, M-1,)(S(m)-\mu(m)), R_k(m))$$

The original image may be reconstructed as a weighted sum of the bases

$$S\backslash D\backslash BA6(\ )(m) = \mu(m) + \backslash f(1,N)\backslash i\backslash su(k = 1, N,)\alpha_k R_k(m)$$

Eliminating the kth basis vector will induce an error on the reconstruction. The average energy of this error is the eigenvalue λ_{kk} from the kth diagonal element of Λ_a

$$E\{(S\backslash D\backslash BA6()(m) - S(m)^2\} = \lambda_{kk}$$

$\lambda_1 = 36.40\%$ $\lambda_2 = 17.67\%$ $\lambda_3 = 9.84\%$

$\lambda_4 = 5.43\%$ $\lambda_5 = 3.94\%$ $\lambda_6 = 3.175\%$

$\lambda_7 = 2.51\%$ $\lambda_8 = 2.09\%$ $\lambda_9 = 1.86\%$
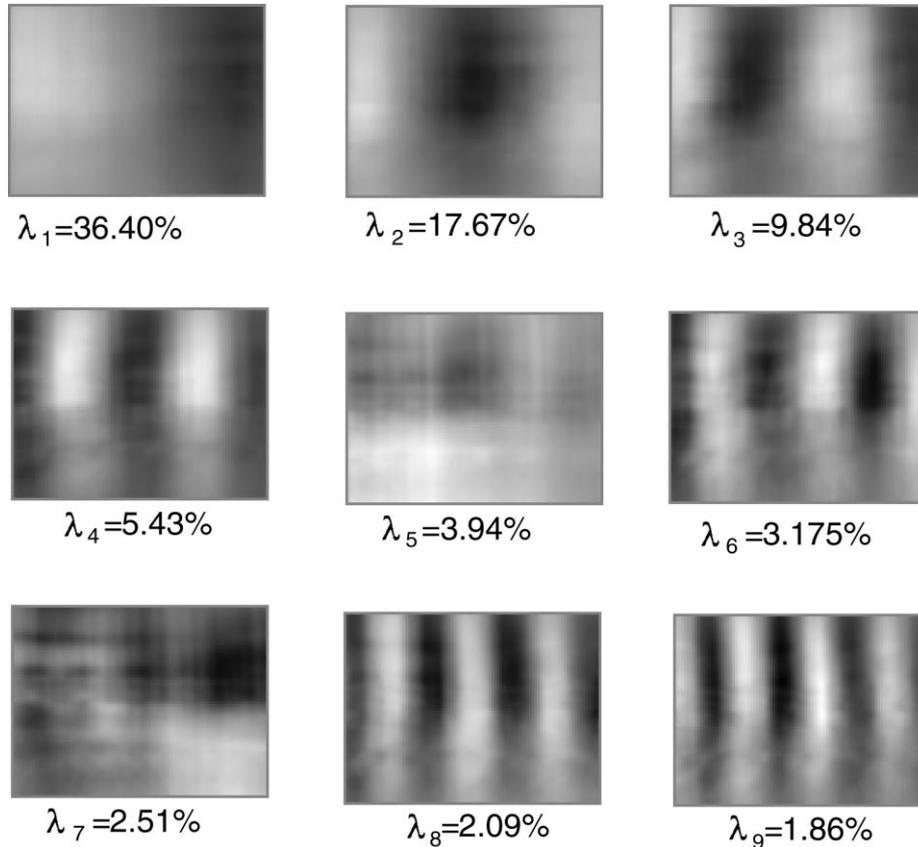
Fig. 2. The first nine principal components and their eigenvalues expressed as percentage of the global sum of the eigenvalues.

Thus the best subset of $R_k(m)$ bases are the K largest eigenvalues of Λ_a such that

$$\sum_{k=K+1,N,} \lambda_{kk} < \text{threshold}$$

Images from the scene which were not in the set $\{S\}$ will provide imperfect reconstructions, with the energy of the error an indication of the similarity of appearance. To minimize this error, we need a relatively dense sampling of position and orientation, and thus a relatively large number of images. The magnitude of this construction error, and thus, the density of required samples can be estimated from the differential properties of the appearance manifold.

Our experience is that the best available algorithms for principal components analysis will provide solutions for covariances matrices up to around 1024×1024. Beyond that size, both the results and the computation times become unpredictable. Thus for image sizes up to 32×32, we can obtain principal components for any number of images using Eqs. (1) and (2). For larger image sizes, we are limited to set of at around 1024 images. Thus, image size (or resolution) is a critical parameter for this approach.

The experiments described below indicate that much of the information for pose estimation is available at relatively low resolutions in the scene. Thus we can use Gaussian pyramid [10,11] to compute low-resolution versions of the images. The Gaussian pyramid provides two important

benefits. Firstly, the reduction in pixels provided by subsampling reduces the number of pixels, M, to a manageable size. Secondly, the Gaussian filtering enables the suppression of high frequencies which ensures a better continuity of the appearance manifold when projected into the eigen space.

We can also reduce the number of sample images, N, by selecting a most representative subset. Sets of similar images are determined by computing a cross correlation between images collected at adjacent positions in the scene, and associating similar images. A representative image can then be extracted from each subset.

As our approach is based on scene appearance, it is dependent on scene illumination. Within the range of intensities over which the camera is linear, variations in illumination intensity can be eliminated by normalizing the image energy for all sample images. Variations in illumination direction require a systematic sampling over the range in variation.

3. A sample representation for the appearance manifold

In order to obtain data for the experiments described below, we constructed four data sets by positioning a steerable camera at a grid of carefully measured positions. In the first data set, (set 1) the camera was moved at 2 cm steps
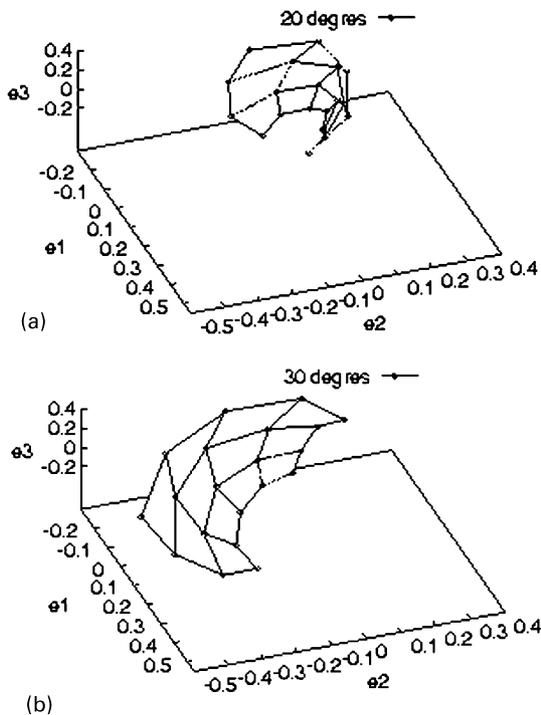
Fig. 3. Projection of a slice of the appearance manifold on the first three principal components from data set 4. The surfaces represent the projection of images taken at 20 and 30°.

along a trajectory in our robotics laboratory with the camera orientation perpendicular to the trajectory. In a second and third data sets (sets 2 and 3), the camera was placed at a grid of positions in the laboratory, separated by 5 cm. In set 2, the camera was oriented along the y axis while in set 3 the camera was oriented at 45° to the grid. In the fourth data set, the camera was placed in a hallway, at a 4×5 grid of positions spaced at 20 cm. In this data set, we programmed the camera to pan over a range of 19 view angles at steps of 10° using the shortest focal length available on the camera ($F = 15$ mm). This produced a data set of 379 images of size 96×72 pixels. Because image set 4 is the most representative of an actual application, we use it to illustrate the appearance manifold. The set up for image set 4 is described in Fig. 1.

We used the covariance from the inner products of images (Eqs. (3) and (5)) to compute the principal components of the 379 images. The first nine principal components for the data obtained in this experiment account for 83% of the total energy Fig. 2. On can notice that the principal components resemble sinusoidal waves. This is consistent with many experiments that show that as the number of images becomes larger and more varied, the principal components of unfiltered images tends to converge toward a Fourier Transform.

The number of principal components required to attain a given percentage of the total sum of eigenvalues tends to increase with the number of pixels. This observation can be readily explained, since high-resolution images contain more details (high frequencies), and thus require a large number of principal components.

The appearance manifold of the scene is represented by a 3D grid (x, y, θ) of discrete samples within a K dimensional basis space. This is illustrated in Fig. 3, which shows the projection from the images taken at the grid of positions for camera angles of 20° (a) and 30° (b). The grid of images is projected onto the first three principal components. The precision of this interpolation depends on smoothness of the manifold with respect to the sample distances Æx, Æy, and Æθ. The smoothness of the surface in Fig. 3 suggests that intermediate images can be synthesized by interpolating over this grid.

In order to apply this approach to real world problems, we need to establish procedures for selecting the image size M, the number of images used to construct the orthogonal basis space, N, the number of basis dimensions to be used, K, and the density of the samples of the manifold, Æx, Æy and Æθ. These parameters are explored in the experiments described in Section 5.

4. Position estimation using appearance

The orthogonal basis provides an index into the appearance manifold that allows an observation to be used as an index to recall view positions with similar appearance. To provide such an index, the K dimensional eigen space is effectively divided into discrete cells. Each cell provides a list of pointers to samples of the appearance manifold which fall within that cell. Such an index may be organized as a tree, with branches allocated dynamically during acquisition of the appearance manifold of the scene [12]. Each branch is organized into a set of bins composed of limited number of samples. If the number of samples in a bin exceeds the limit, then the bin may be divided. Each bin points to a set of bins for the next higher dimension. The bins for the final (Kth) dimension point to a list of samples on the appearance manifold. Thus, the leaves of the tree provide a list of position and orientations with the same appearance.

In real scenes, it is not uncommon for some view angles to produce uniform images. Because of energy normalization, uniform images project to an appearance vector which is nearly null. To avoid problems with such regions, appearance vectors for which the total energy is close to zero are not included in the indexing structure and are not used for indexing. Similarly, whenever a leaf in the indexing structure contains an excessive number of nearly identical appearance vectors, the leaf is marked as unusable for indexing.

The position and orientation of the camera may be estimated by projecting an unknown image, $S(m)$, into the appearance space to give an appearance vector, α_k. A form of branch and bound is used to assemble lists of
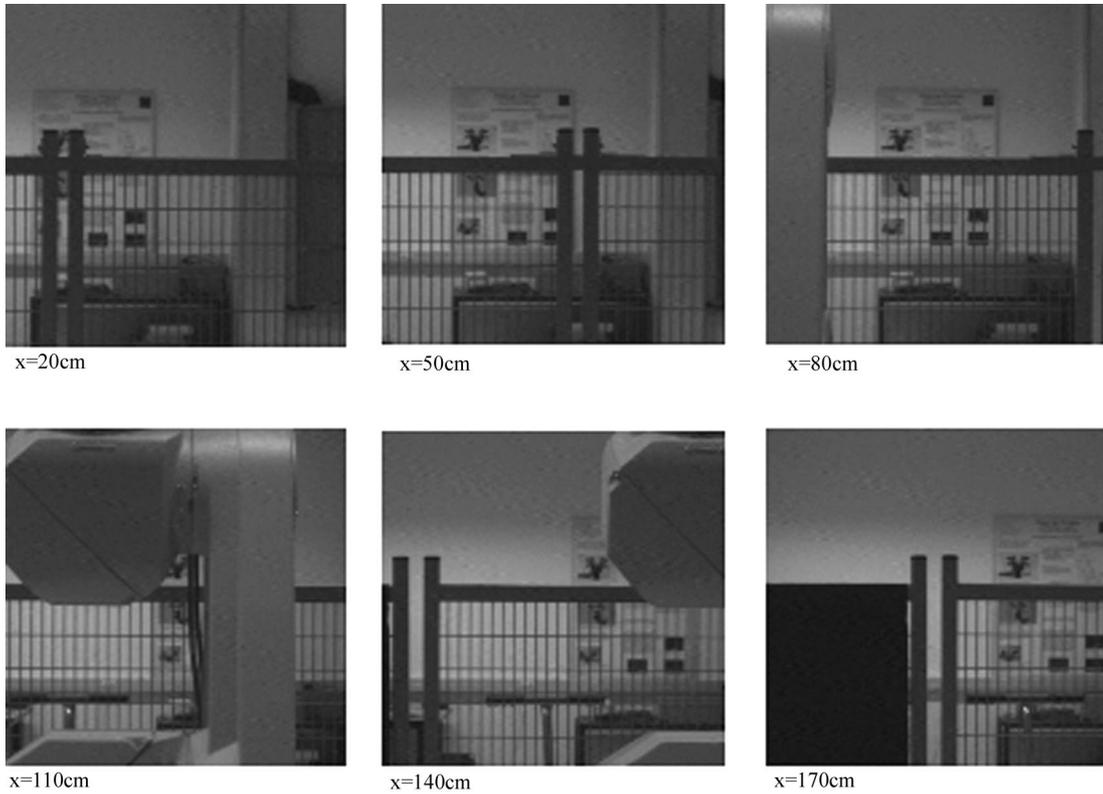
x=20cm
x=50cm
x=80cm

x=110cm
x=140cm
x=170cm

Fig. 4. Six images from sequence of 100 images obtained every at 2 cm intervals with the camera looking perpendicular to the path.

hypothesis with similar appearance, as measured by the distance between the appearance vectors. The distance metric is defined by normalizing each dimension by the covariance of values observed during acquisition of the appearance manifold. The threshold for similarity, ϵ_{sim}, is a parameter which is tuned during system set-up.

It is not uncommon for regions of an appearance manifold to have similar appearance. The list of hypotheses can be greatly reduced by grouping hypotheses into sets based on connectivity in the appearance manifold. Connected sets of

hypotheses can be used to fit a parametric function to the manifold, and a more precise hypothesis can be computed by projecting the observed image onto this function. In Section 5 we compare the use of cubic and linear interpolation for such estimation.

The list of grouped hypotheses are sorted based on the distance between the projection and the pre-learned appearance vectors. The best hypothesis can be selected by a tracking process based on previous observations [13,14]. We believe that this selection process can be greatly enhanced

hdng0

$\lambda_1 = 24.32\%$
$\lambda_2 = 22.22\%$
ard $\lambda_3 = 9.75\%$
$\lambda_4 = 6.60\%$
$\lambda_5 = 6.60\%$

hdng0

$\lambda_6 = 5.46\%$
$\lambda_7 = 4.65\%$
$\lambda_8 = 4.04\%$
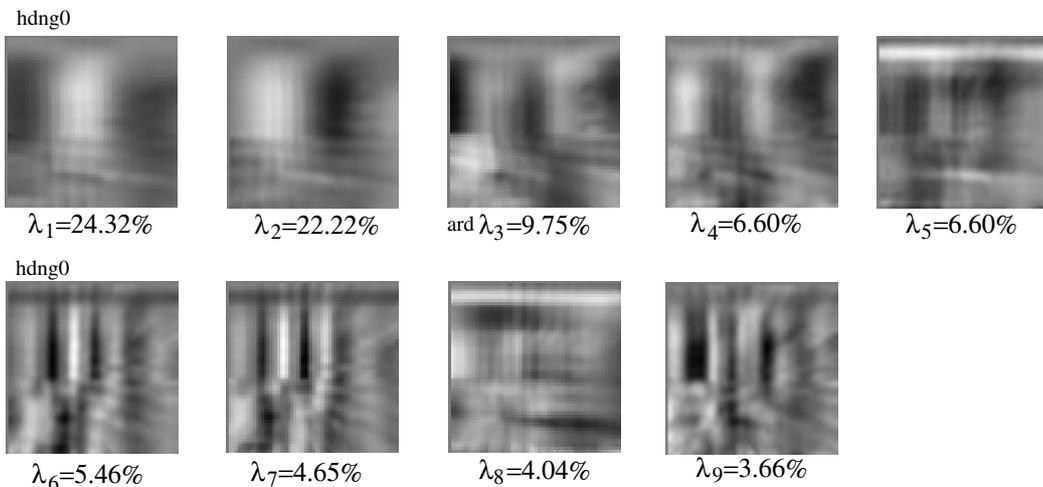$\lambda_9 = 3.66\%$

Fig. 5. The first nine principal components and their principal values(expressed as a percentage of the total) from the images in Fig. 4.
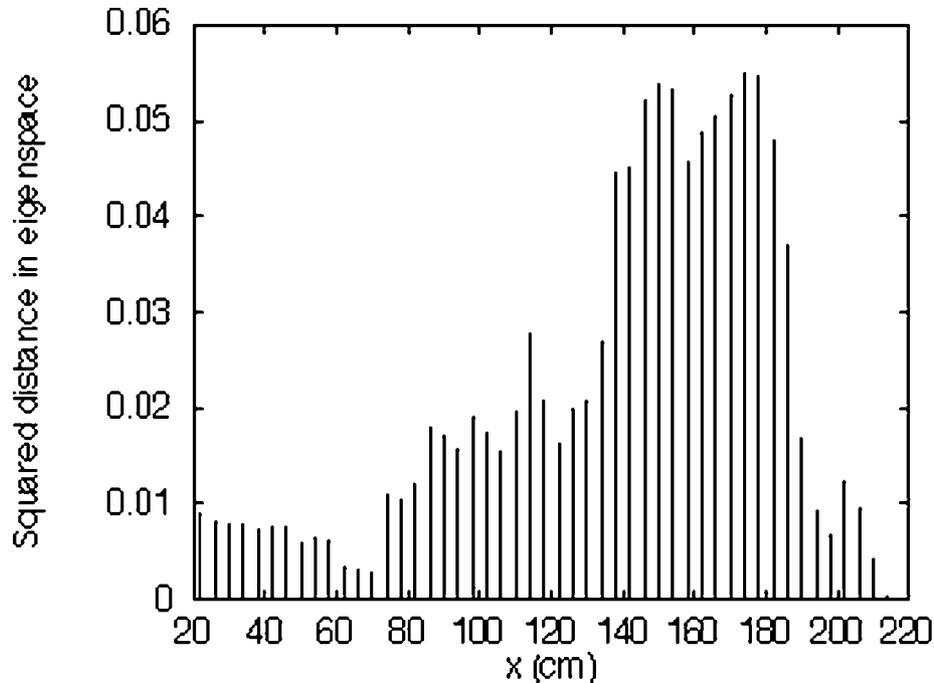
Fig. 6. This figure shows the distance in appearance space for adjacent images as a function of the position along the path in data set 1.

by integrating the results into a tracking process based on odometry. However, we do not use odometry in the experiments described later.

5. Experimental investigation of parameters

As stated above, using an appearance manifold for position estimation requires establishing several parameters. The following experiments explore the influence of these parameters on the precision of the resulting estimation using the four data sets described earlier.

5.1. Localization along a path

Data set 1 is composed of 100 images of size 64×64
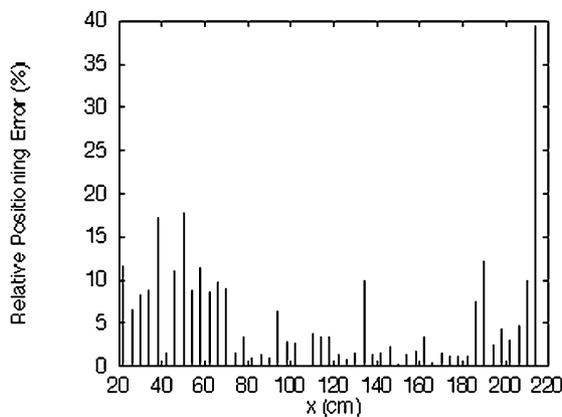
Fig. 7. Shows the relative position error for manifolds constructed from images taken at positions spaced by 20–200 cm.

pixels collected at steps of 2 cm along a path in our robotics laboratory. The camera was set to look perpendicular to the path at a scene composed of a poster on a wall, and a robot arm protected by a protective barrier with the appearance of a grid. The orientation was chosen to be perpendicular to the motion in order to obtain the strongest visual variations between successive images. Nine images from this sequence are shown in Fig. 4. The first nine principal components from this image set are shown in Fig. 5.

The similarity between images on the manifold can be measured using an L_2 distance metric. Fig. 6 shows the squared distance in principal component space between adjacent images in the sequence as a function of position along the path. One can observe that at the beginning and end of the sequence, appearance changes slowly, while at the center, appearance changes more rapidly. This histogram suggests that the matching tolerance, ϵ_{sim}, must be at least 0.055 for this data set.

The average accuracy of a position estimation depends on the sampling density of the scene as well as on the environment structure. A first experiment investigates the effects of varying the density of view points used to build the manifold. Appearance manifolds were constructed by projecting subsets of the images onto the first nine principal components. Images not used in constructing the manifold were projected onto the manifold in order to estimate the position along the path. Position along the path was estimated using a linear interpolation between the nearest points on the manifold. Fig. 7 shows the error in estimated position compared to the step size in meters between samples used to build the manifold. This corresponds to the relative position error as a function of the distance viewing positions of the images
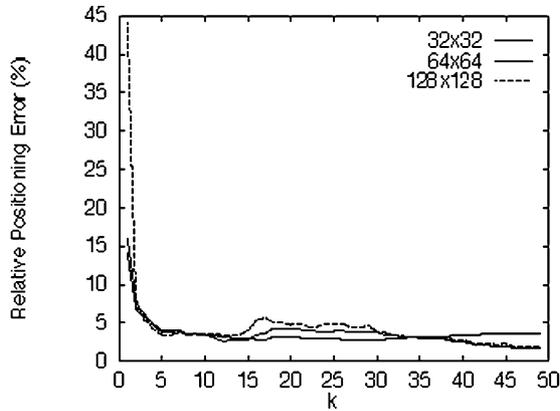
Fig. 8. Shows the relative position error as a function of the number of principal components used for representing the appearance manifold. The data is displayed for 32×32, 64×64 and 128×128 images.

used to build the manifold. We observe that up to 80 cm and after 200 cm, the relative positioning error is high since the squared distance is small. Furthermore, when a new scene feature appears in the image, the error grows. This corresponds to the peaks recorded around 130 and 190 cm, which can be explained by the large difference of distance between points lying in the neighborhood of these peaks as shown in Fig. 6. Our conclusion is that linear interpolation makes it possible to greatly reduce the density of samples required in building the manifold with only minimal loss in precision.

5.2. Image size and number of dimensions

As illustrated in Section 3, the image size and number of bases are crucial questions for the usability of such an approach for practical applications. These issues are addressed by a systematic comparison of position error over different parameters. The results of such a comparison are shown in Fig. 8. This figure compares position error as a function of the number of dimensions for different size images for data set 1. Although the best results were obtained for k varying from 10–15 dimensions, these results were only marginally better than those obtained with 4 dimensions. Increasing either the image resolution or the
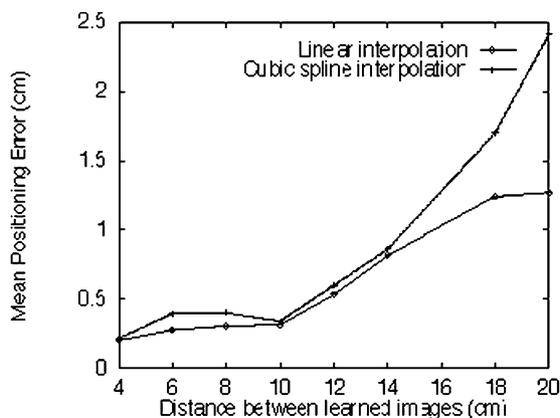
number of dimensions provides only marginal improvement in position estimation accuracy at a substantial increase in computing cost.

The conclusion from this experiment (and others) is that reasonably good position estimation results can be obtained from low resolution images represented in a low dimensional appearance space. Position estimation tends to rely on visual details which are at low resolution captured in images of 32×32 (or even 16×16 as observed in later experiments). Furthermore, using low-resolution images allows the principal components to be computed using Eqs. (1) and (2), which makes it possible to greatly increase the number of images included in the covariance matrix. Increasing the number of dimensions beyond 5 provides only a very marginal improvement in position estimation, at a large increasing in computing and memory costs.

5.3. Linear versus cubic interpolation

To evaluate the interpolation quality, every 10th image from data set 1 was used to build an appearance manifold, while the others were used for testing. We compared the use of linear and cubic interpolation between grid points for position estimation. The results, presented in Fig. 9, show that linear interpolation provided a more accurate interpolation than a cubic spline. The reason is illustrated in Fig. 10. When projected to a principal component space, the appearance manifold contains regions which are relatively smooth bounded by regions with a very sharp curvature. Both cubic spline and linear interpolation perform reasonably well over the smooth regions. However, a cubic spline tends to overshoot at the boundaries where curvature in the manifold is sharp.

5.4. Position estimation using a 2D manifold

To test position estimation in a 2D grid, two data sets (sets 2 and 3) were constructed from 400 images each of size 64×64 pixels acquired with the camera positioned at a 20×20 grid of samples separated by 5 cm. The scene is the same as in data set 1. In data set 2, the images were acquired with the camera oriented along they axis, with the result that the y axis is a change in depth in the scene. In data set 3, the camera turned to 45°. Fig. 11 shows the images from the corners of the gird from data set 2.

For both data sets 2 and 3, the first 18 principal components represent 95% of the variance. The projection of the appearance manifold for data set 3, plotted on the first three principal components is represented in Fig. 11.

An appearance manifold was constructed using a 10×10 subset of the data set 2, with the view positions separated by 10 cm as shown in Fig. 12. Test images taken from positions off-set from the learned images by 5 cm in x and y were used to estimate position, using linear interpolation. The result is shown in Fig. 13. The positions used to construct the manifold are shown in the grid of squares. The true position for each test image is shown by an "o" while the estimated

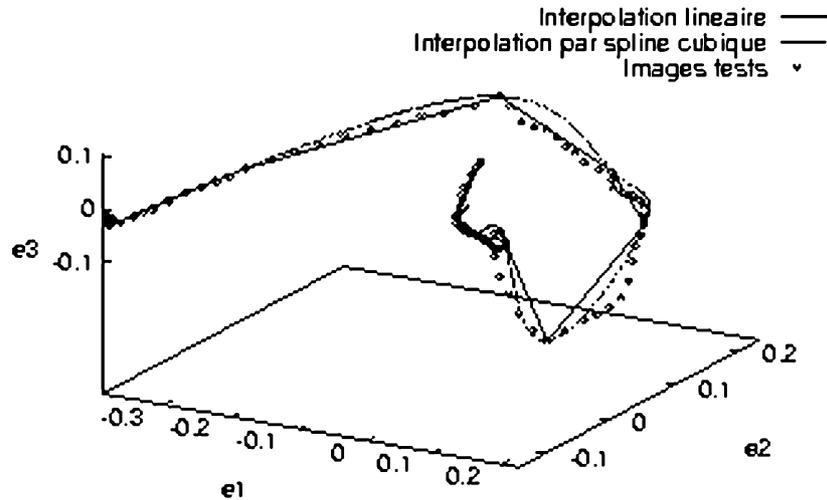Fig. 9. Comparison of linear and cubic interpolation.

Fig. 10. Comparison of manifolds constructed using linear and cubic interpolation using every 10th point as a control point. The actual appearance data are displayed as " * ". The 1D appearance manifold is displayed for the first three principal components (e_1, e_2, e_3) shown in Fig. 5.

position is shown by an "x". As expected, the position estimate is very precise for displacements perpendicular to the camera direction (Fig. 14a), and less precise for displacements in the camera direction Fig. 14b. The error ellipse constructed from the covariance of the position estimation errors is shown in Fig. 15a. The experiment was repeated with data set 3 giving an error covariance aligned with the camera direction, as shown in Fig. 15b. From these results we can conclude that the error of position estimation is highly correlated with camera orientation.

5.5. Position and orientation estimation using a 3D manifold

To test the use of a 3D manifold defined over x, y, and θ, an appearance manifold was constructed from the 379 images in data set 4 (shown in Fig. 1). A test set of 25 images were made from intermediate positions offset from the learned positions by 10 cm in x and y. The camera for the test set was oriented at 45°. The results are shown in Fig. 16. Fig. 17 shows histograms for the
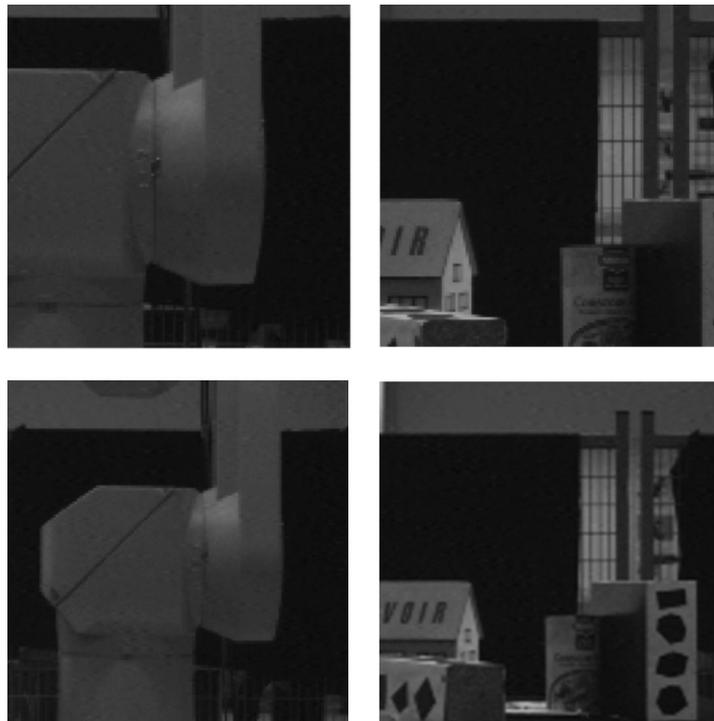

Fig. 11. Images at the extreme corners of data set 2. The y axis corresponds to a change in scale of the scene.
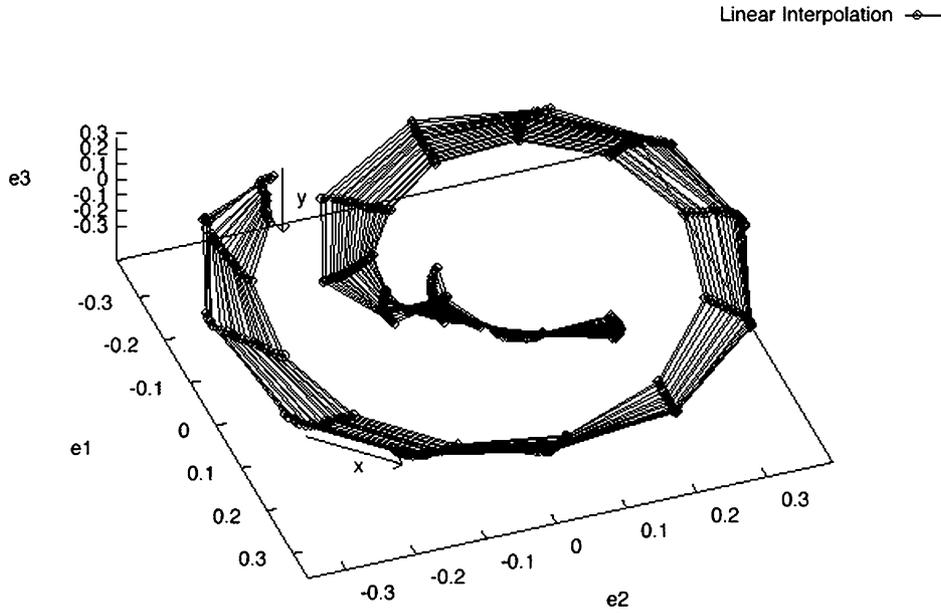
Linear Interpolation

Fig. 12. The appearance manifold from data set 2 shown in Fig. 11.

errors in the *x* and *y* directions. These results demonstrate that the position and orientation of a mobile robot can be estimated using linear interpolation of over a 3D appearance manifold.

6. Conclusions

A discrete sampling of images from an environment can be used to construct a discrete approximation of an appearance manifold for position estimation. Projecting the appearance manifold onto the first few principal components from a representative subset of the images provides a method for fast indexing for position estimation. Linear interpolation can be used to provide a gain of about a factor of 10 in the resulting position estimation. Position estimation can easily be computed at video rates using conventional micro-processors.

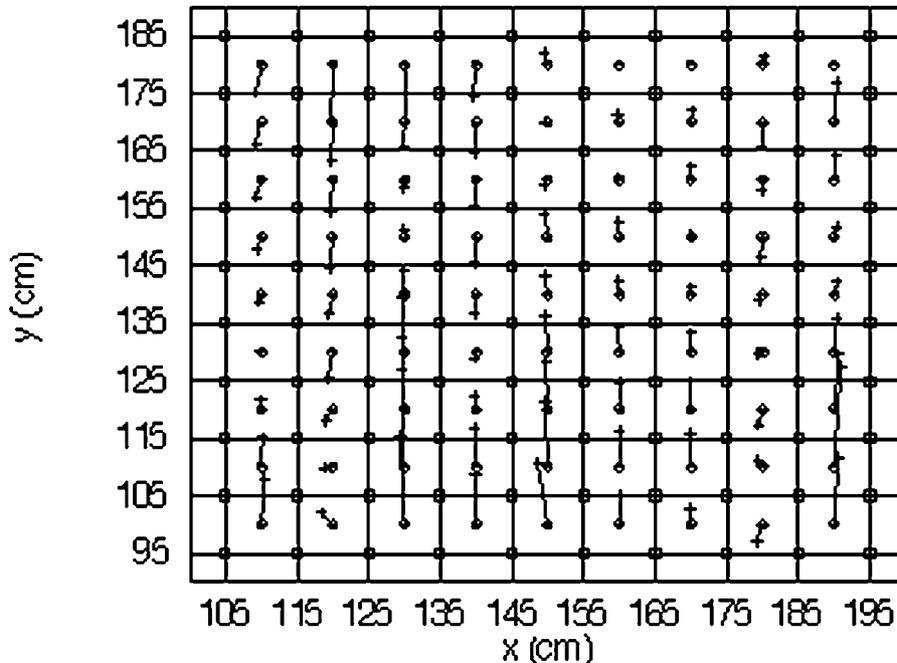Our experiments show that this approach can work with an appearance space constructed from a small

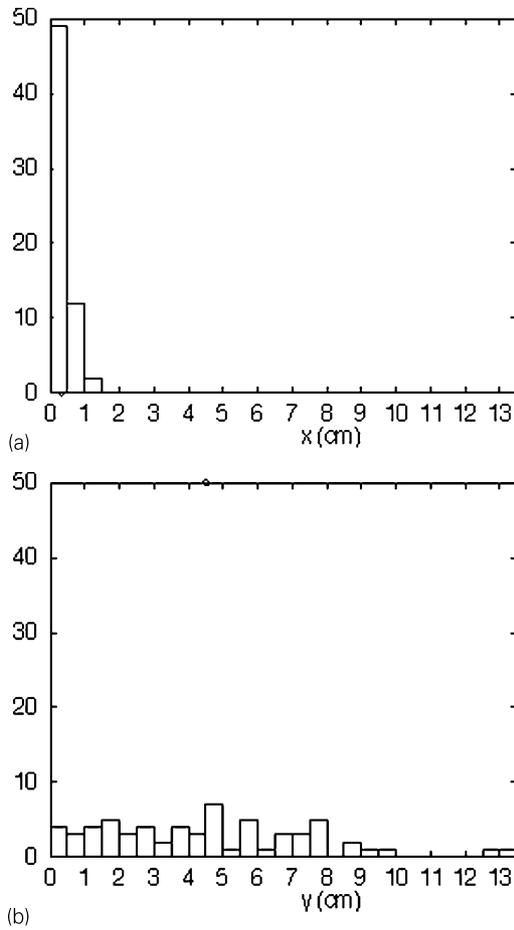Fig. 13. Position errors obtained for grid of positions in data set 2.

(a)

(b)

Fig. 14. Histogram of position errors in the *x* and *y* directions for data set 2.

(a)

(b)
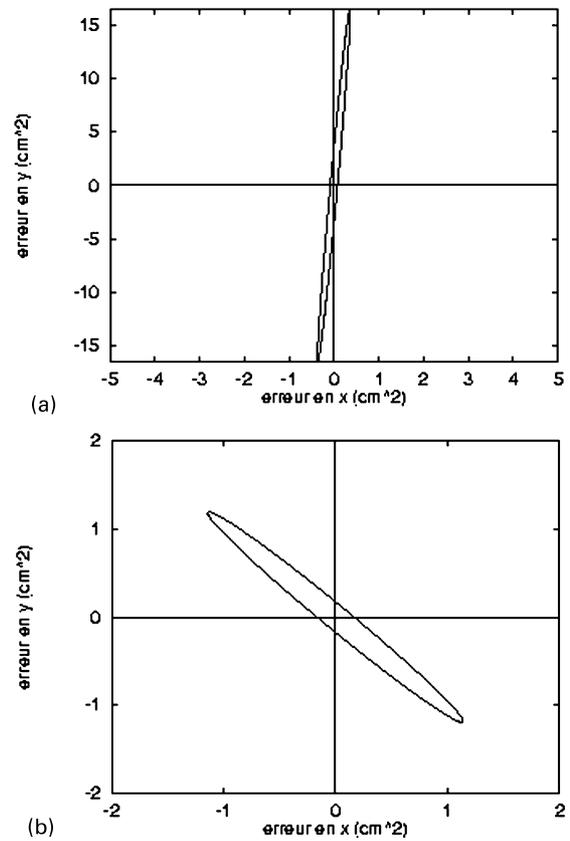
Fig. 15. (a) Shows the error ellipse (1-σ) for position estimation errors for position estimation experiment described in Fig. 13 (data set 2). The error ellipse is (nearly) oriented along the camera direction. (b) Shows the error ellipse for the same experiment using data set 3, with the camera turned 45°.

number of dimensions using relatively low resolution images. Using small images allows the principal components to be computed directly on the covariance of pixels, thus allowing a very large number of images to be used in constructing the principal components. This allows a more complete coverage of the appearance space of a scene. Using a small number of principal components is useful for reducing the memory requirements, but is not a critical requirement. Indeed, the number of dimensions can be selected dynamically to adapt to the diversity of the scene. The experiments described here deal with position estimation within a limited region of the environment (for example a room or a corridor). A large-scale map can be built from a collection of such local maps.

The experiments described in this paper are very exploratory and the method is characterized by some limitations. For example, the problem of position estimation in a dynamically changing environment is not addressed. We believe that the method can be made more robust to moving obstacles by use of a local method based on Gaussian derivatives such as used by Colin de Verdiere [12].
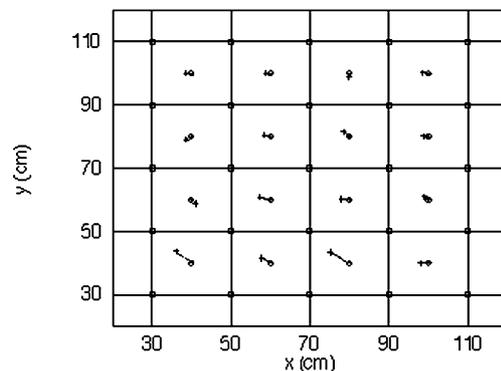

Fig. 16. Position estimation errors using a 3D manifold constructed from data set 4 (Fig. 1). The test images were acquired off-set by 10 cm in *x* and *y*, with an orientation of 45°. The true position for reach test image is shown as a "o". The position estimated by linear interpolation is shown by a "x".
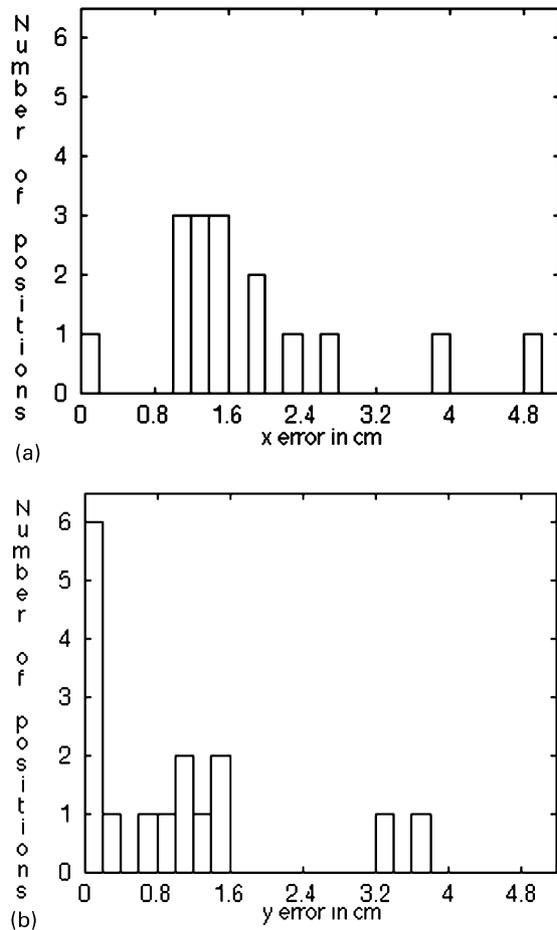
s

Fig. 17. Histogram of position errors in *x* and *y* directions for experiment with data set 4 (hallway scene). The appearance manifold was sampled at 20 cm steps in position and 10° steps in orientation. Position was estimated using linear interpolation. The test data was made with the camera oriented at 45°.

References

[1] M. Kirby, L. Sirovitch, Application of the Karhunen–Loeve procedure for the characterization of human faces, PAMI 12 (1) (1990) 103–108.

[2] M.A. Turk, A.P. Pentland, Eigen faces for recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.

[3] H. Murase, S.K. Nayar, Visual learning and recognition of 3D objects from appearance, IJCV 14 (1995) 5–24.

[4] M.J. Black, A.D. Jepson, Eigen tracking: robust matching and tracking of articulated objects using a view-based representation, ECCV (1996) 329–342.

[5] S. McKenna, S. Gong, Real-time face pose estimation, International Journal on Real Time Imaging (Special Issue on Real-time Visual Monitoring and Inspection) 4 (1998) 333–347.

[6] F. Wallner, Position estimation for a mobile robot from principal components of laser range data, doctoral dissertation, INP Grenoble, Grenoble, France, 1997.

[7] J.L. Crowley, F. Wallner, B. Schiele, Position estimation using principal components of range data, 1998 IEEE Conference on Robotics and Automation, Leuven, May 1998.

[8] M. Lhuillier, L. Quan, Image interpolation by joint view triangulation, Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA, vol. 2, 1999, pp. 139–145.

[9] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C: the Art of Scientific Computing, Cambridge University Press, Cambridge, MA, 1992.

10 J.L. Crowley, R.M. Stern, Fast computation of the difference of low-pass transform, IEEE Transactions on PAMI, PAMI 6 (2) (1984).

[11] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Transactions on Communications 31 (4) (1983).

[12] V. Colin de Verdiere, J.L. Crowley, Local appearance space for recognition of navigation landmarks, Robotics and Autonomous Systems 31 (1/2) (2000) 61–70.

[13] J.L. Crowley, World modeling and position estimation for a mobile robot using ultrasonic ranging, IEEE Conference on Robotics and Automation, Scottsdale, AZ, May 1989.

[14] J.L. Crowley, P. Reignier, Asynchronous control of rotation and translation for a robot vehicle, Robotics and Autonomous Systems 10 (1) (1993).