Published in proceedings of CVWW'04, pages 205-214, 2004.

Appearance-Based Localization using CCA*

Danijel Skočaj and Aleš Leonardis
University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, SI-1001 Slovenia
e-mail: danijel.skocaj@fri.uni-lj.si

Abstract

In this paper we present an appearance-based approach to mobile robot localization based on Canonical Correlation Analysis. The main idea is to learn the relation between the appearances of the environment from a number of training locations and coordinates of these locations using CCA and then to use this knowledge to estimate the position of the robot in the localization stage. We present results of several experiments, which show that this approach is faster and less demanding in terms of space than traditional PCA-based approach, however in its standard form it yields in general inferior localization results.

1 Introduction

Self-localization is one of the crucial capabilities of a mobile robot that enables autonomous navigation in the environment. To achieve this, the robot has to collect data about the environment and use it to determine its current position. Many types of sensors can be used to collect the data (sonar, laser beam, infra-red sensors, etc.), however the most natural and non-intrusive are visual sensors. Visual information is very rich, and as such is also the main source of information for self-localization of humans and most of the animals.

A plethora of approaches to mobile robot localization based on visual input has been proposed in the past (for a comprehensive survey see [2]). One of those is appearance-based approach, where robot localization is based directly on the acquired (usually panoramic) 2-D images. It can be realized using subspace methods, which have been widely used for building representations of objects or scenes from their appearances. Among them, the most known is Principal Component Analysis [3]. In this work, however, we focus on Canonical Correlation Analysis (CCA), which is specifically tailored for regression tasks, such as estimation of the position.

*This work was supported in part by the EU project *Cognitive Vision Systems - CogVis* (IST-2000-2937), the grants funded by the Ministry of Education, Science and Sport of Republic of Slovenia: Research Program *Computer Vision-1539-506* and *SLO-A/07*, and by the Federal Ministry for Education, Science and Culture of Austria under the *CONEX* program.

CCA [1, 5] is a supervised method, which relates two sets of observations, one set being composed of training images and the other set of the corresponding measurements (e.g., orientations or positions of the robot; see Fig. 1). CCA finds pairs of directions that yield maximum correlation between the projections of input vectors. We can then perform linear regression on the obtained projections (canonical coefficients). Later, we can estimate the orientation or position of the robot by using canonical coefficients obtained from a novel image of the current scene. This approach is significantly simpler and faster then the PCA-based localization, where all the representations of the training images have to be stored and where the position of the robot is determined by searching for the representation, which is the closest to the representation of the novel image [3].

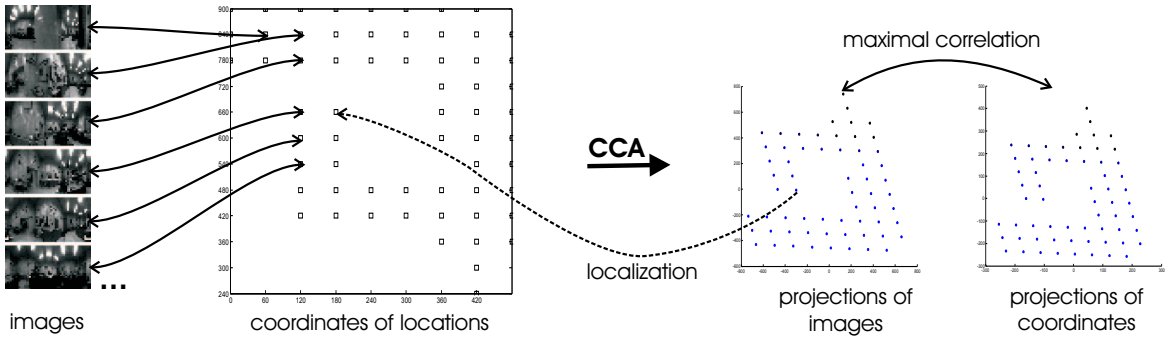


Figure 1: Principle of CCA-based localization.

In CCA, the number of obtained canonical correlation vectors is bounded by the lower dimension of the observations. Since the second set of observations (orientation, position of the robot) is usually low-dimensional, CCA yields only a few canonical correlation vectors. On one hand this is advantageous, since it speeds up the processing, while on the other hand the representation holds less information. The main question is if the information contained in two or three canonical correlation vectors suffices for reliable self-localization. We will try to answer this question by performing several experiments and comparing the results with the traditional PCA-based method.

2 Basic CCA and dual formulation

We will first present the basic concepts of canonical correlation analysis. We will review the standard derivation of CCA and its dual formulation, which enables employment of CCA on high-dimensional input vectors, such as images. We will follow the presentation and derivation given in [5].

Given N pairs of mean-normalized observations $(\hat{\mathbf{x}}_i, \hat{\mathbf{y}}_i)\ ,\ i = 1 \ldots N$, aligned in the data matrices $\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_N] \in \mathbb{R}^{p \times N}$ and $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_N] \in \mathbb{R}^{q \times N}$, CCA finds pairs of directions $\mathbf{w_x} \in \mathbb{R}^p$ and $\mathbf{w_y} \in \mathbb{R}^q$ that maximize the correlation between the projections

$\mathbf{w_x}^\top \hat{\mathbf{x}}_i$ and $\mathbf{w_y}^\top \hat{\mathbf{y}}_i$. CCA maximizes the function

$$\rho = \frac{\mathbf{w_x}^\top \mathbf{C_{xy}} \mathbf{w_y}}{\sqrt{\mathbf{w_x}^\top \mathbf{C_{xx}} \mathbf{w_x} \mathbf{w_y}^\top \mathbf{C_{yy}} \mathbf{w_y}}} , \tag{1}$$

where $\mathbf{C_{xx}} = \frac{1}{N}\hat{\mathbf{X}}\hat{\mathbf{X}}^\top \in \mathbb{R}^{p\times p}$ and $\mathbf{C_{yy}} = \frac{1}{N}\hat{\mathbf{Y}}\hat{\mathbf{Y}}^\top \in \mathbb{R}^{q\times q}$ are within-set covariance matrices and $\mathbf{C_{xy}} = \frac{1}{N}\hat{\mathbf{X}}\hat{\mathbf{Y}}^\top \in \mathbb{R}^{p\times q}$ and $\mathbf{C_{yx}} = \frac{1}{N}\hat{\mathbf{Y}}\hat{\mathbf{X}}^\top \in \mathbb{R}^{q\times p}$ are between-set covariance matrices of input data.

We will refer to the extremum points $\mathbf{w_x}^*, \mathbf{w_y}^*$ of (1) as *canonical correlation vectors*, whereas the projections of the input observations onto the canonical vectors will be referred to as *canonical correlation coefficients*. The extremum values $\rho^* = \rho(\mathbf{w_x}^*, \mathbf{w_y}^*)$ are the canonical correlations and are as large as possible.

Several approaches to the maximization of (1) have been proposed. Here we present a formulation with Rayleigh quotient [1]. Let us arrange covariance matrices in block matrices $\mathbf{A} \in \mathbb{R}^{(p+q)\times(p+q)}$ and $\mathbf{B} \in \mathbb{R}^{(p+q)\times(p+q)}$:

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & \mathbf{0} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{C_{xx}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_{yy}} \end{bmatrix} \tag{2}$$

and concatenate $\mathbf{w_x}$ and $\mathbf{w_y}$ into vector $\mathbf{w} = \begin{bmatrix} \mathbf{w_x} \\ \mathbf{w_y} \end{bmatrix} \in \mathbb{R}^{p+q}$. It can be shown [1] that the extremum points $\mathbf{w_x}^*, \mathbf{w_y}^*$ of ρ (Eq. 1) coincide with the stationary points $\mathbf{w}^* = \begin{bmatrix} \mathbf{w_x}^{*\top} & \mathbf{w_y}^{*\top} \end{bmatrix}^\top$ of the Rayleigh quotient

$$r = \frac{\mathbf{w}^\top \mathbf{A} \mathbf{w}}{\mathbf{w}^\top \mathbf{B} \mathbf{w}} . \tag{3}$$

According to Generalized spectral theorem, these extremum points can be obtained as the eigenvectors of the corresponding generalized eigenproblem:

$$\mathbf{A}\mathbf{w} = \lambda \mathbf{B}\mathbf{w} . \tag{4}$$

Since the matrices $\mathbf{A}$ and $\mathbf{B}$ are of the dimension $(p+q)\times(p+q)$, their size directly depends on the size of the observations. Since vectors representing input images are usually very high-dimensional, the size of the covariance matrix $\mathbf{C_{xx}}$ and consequently matrices $\mathbf{A}$ and $\mathbf{B}$ becomes prohibitively large. If the number of the observations N is smaller than the size of the images p, thus $p \gg N \gg q$, the following dual formulation of CCA [5] can significantly improve the efficiency of the calculation of canonical vectors.

Since every solution component vector $\mathbf{w_x}^*$ lies in the span of the training data i.e., $\mathbf{w_x}^* \in span(\hat{\mathbf{X}})$ (for proof see [5]), there exists a vector $\mathbf{w'_x}^* \in \mathbb{R}^N$, so that

$$\mathbf{w_x}^* = \hat{\mathbf{X}} \mathbf{w'_x}^* . \tag{5}$$

Considering this, it can be shown that we can set up a dual problem to (4):

$$\mathbf{A}'\mathbf{w}' = \lambda \mathbf{B}'\mathbf{w}' , \tag{6}$$

$$\text{where } \mathbf{w}' = \begin{bmatrix} \mathbf{w}'_{\mathbf{x}} \\ \mathbf{w}_{\mathbf{y}} \end{bmatrix} \in \mathbb{R}^{N+q}, \quad \mathbf{A}' = \begin{bmatrix} \mathbf{0} & \mathbf{C}'_{\mathbf{xx}}\hat{\mathbf{Y}}^{\top} \\ \hat{\mathbf{Y}}\mathbf{C}'_{\mathbf{xx}} & \mathbf{0} \end{bmatrix}, \quad \mathbf{B}' = \begin{bmatrix} {\mathbf{C}'_{\mathbf{xx}}}^{2} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\mathbf{yy}} \end{bmatrix}. \quad (7)$$

Here, $\mathbf{C}'_{\mathbf{xx}} = \hat{\mathbf{X}}^{\top}\hat{\mathbf{X}} \in \mathbb{R}^{N \times N}$ is the inner-product matrix, which is significantly smaller than the covariance matrix $\mathbf{C}_{\mathbf{xx}} \in \mathbb{R}^{p \times p}$ when $N \ll p$. Consequently, also the matrices $\mathbf{A}' \in \mathbb{R}^{(N+q)\times(N+q)}$ and $\mathbf{B}' \in \mathbb{R}^{(N+q)\times(N+q)}$ are smaller, which alleviates the problem of high-dimensional data. After the generalized eigenvalue problem (6) is solved, the canonical correlation vectors ${\mathbf{w}_{\mathbf{x}}}^*$ are obtained by projecting the solutions ${\mathbf{w}'_{\mathbf{x}}}^*$ into the image space using (5).

3 CCA for self-localization

Appearance-based localization is a twofold procedure: in the *learning stage*, several panoramic images are acquired from different positions, which together form a good depiction of the environment. For each training image $\mathbf{x}_i$ the corresponding position $\mathbf{y}_i$ is known (e.g., two-dimensional vector indicating two offsets from the starting position). Then the CCA (or its dual formulation) is applied to these two sets of training data. As a result the canonical correlation vectors of both input sets are obtained. Then we project all training images to their CCA vectors (aligned in the matrix $\mathbf{W}_{\mathbf{x}} \in \mathbb{R}^{p \times q}$) and obtain vectors of canonical correlation coefficients aligned in the matrix $\mathbf{V}_{\mathbf{x}} \in \mathbb{R}^{q \times N}$, thus $\mathbf{V}_{\mathbf{x}} = {\mathbf{W}_{\mathbf{x}}}^{\top}\hat{\mathbf{X}}$. We estimate a linear mapping $\mathbf{F}$ from these q(usually two)-dimensional canonical correlation coefficients to the corresponding $\mathbf{y}_i$ using the least squares minimization method, i.e., $\mathbf{F} = \hat{\mathbf{Y}}{\mathbf{V}_{\mathbf{x}}}^{\dagger}$. The q canonical correlation vectors in $\mathbf{W}_{\mathbf{x}}$ and the matrix $\mathbf{F} \in \mathbb{R}^{q \times q}$ form the model of the environment and are the only data we have to keep. Therefore, we need to store $qp + qq$ elements (or $(q+1)p + (q+1)q$ if the input data is not mean-centered, therefore we have to keep the mean vectors also).

Later, in the *localization stage*, all we have to do to estimate the location of the robot ($\hat{\mathbf{y}}$) is to capture the panoramic image $\hat{\mathbf{x}}$, project it onto the canonical correlation vectors ($\mathbf{v}_{\mathbf{x}} = {\mathbf{W}_{\mathbf{x}}}^{\top}\mathbf{x}$) and map the obtained canonical correlation coefficient vector $\mathbf{v}_{\mathbf{x}}$ using the mapping function $\mathbf{F}$, thus $\hat{\mathbf{y}} = \mathbf{F}\mathbf{v}_{\mathbf{x}}$. Therefore, we need to perform only $qp + q^3$ simple multiplications.

It is evident that the storage and time requirements of this approach are significantly smaller than in the traditional PCA-based approach [3]. In the latter case, we have to store k principal vectors ($k \gg q$) and N k-dimensional vectors of principal components (or even more if we want to increase the accuracy of the results by interpolating between the coefficients of the training images), thus at least $kp + Nk$ elements. Also the time complexity of the localization stage is significantly higher in the PCA-based approach, since after projecting the novel image onto the k principal vectors (requiring kp multiplications), we have to search for the closest point in the k-dimensional principal subspace.

Besides the space and time complexity, the CCA has also an additional advantage over the PCA-approach. The domain of the results is continuous, while in the case of PCA approach it is discrete, limited to the training data and interpolations. From these points of view, CCA-based approach is better suited for mobile robot localization. However the main question is how good the localization is. Are only q canonical correlation vectors enough for reliable localization? We will answer this question in the next section.

4 Experiments and discussion

Most of the experiments presented in this section were performed on a sequence of images taken in CMP lab at the Prague Technical University (six of them are shown in Fig. 2). The plots in Figs. 3 and 5 depict a map of the lab. The training images were taken from the positions marked with squares. The actual path of the robot in the test stage (ground truth) is indicated with green, while the estimated path is marked with blue points and lines. The error at each test location is depicted as a red dashed line. For most of the experiments such graphical qualitative results are presented, as well as quantitative results in terms of mean localization error (the average distance between the actual and the estimated locations). In these experiments we used 62 training and 100 test panoramic images unwarped and resized to 26×50 pixels.

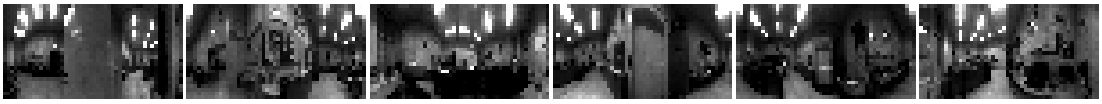


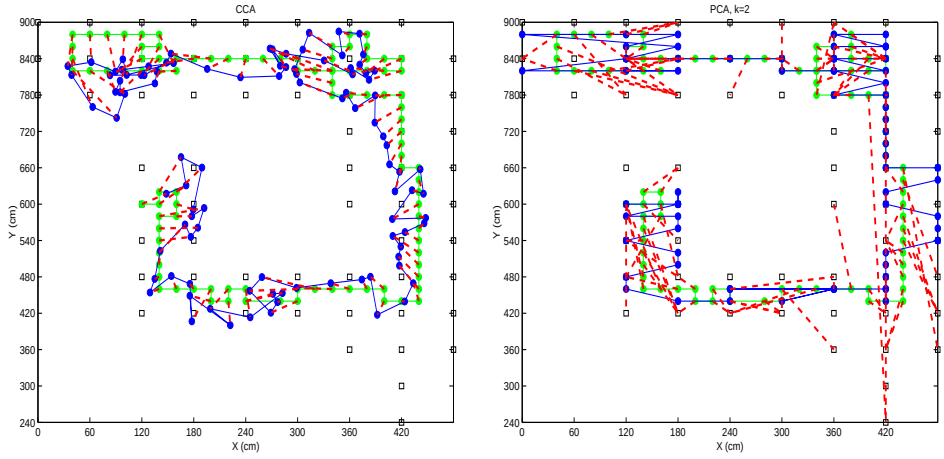
Figure 2: Six training images from CMP sequence.

4.1 2DOF

First we present the results of the localization in two degrees of freedom. The goal was to determine the two offsets from the origin of the map. We assumed that the robot was able to determine its orientation (e.g., by using compass or some other technique [3, 6]) thus it was always rotated in the same direction and we did not have to model all possible rotations. The results for various approaches are presented in Fig. 3 and Table 1(a).

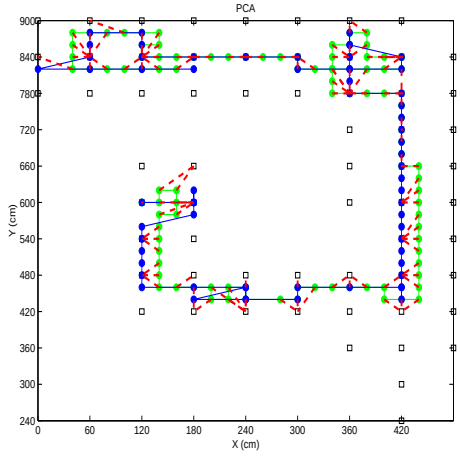
One can observe that the described CCA-based approach yields 34.77 cm average localization error. All hundred test locations together were processed in 0.02 seconds only. Two canonical correlation vectors were used to model the environment and 3906 matrix elements were used altogether to store the model. The PCA-based approach took significantly longer and produced inferior results when only two principal vectors were used (error 77.37 cm). However, in this case we were able to use a larger subspace. When we stored 10 principal vectors, the results were significantly better (error 26.16), and when we used the interpolation-based approach [3], the results were even further improved, arriving at very good 13.04 cm of error. The price we had to pay were large storage requirements and very long processing time (2.77 seconds for hundred images - this processing time could be decreased by implementing a better search algorithm, however it would be still significantly larger than in the case of CCA-based localization.)

From this we can conclude that CCA-based approach is very fast, however produces inferior results than PCA-based approach in general. The main reason is that we can use two canonical correlation vectors only, since the dimension of the location vectors is two. One solution would be to increase this dimension by using kernelized version of CCA [5]. KCCA yields almost identical results as PCA for equal dimension of the subspace. However, since

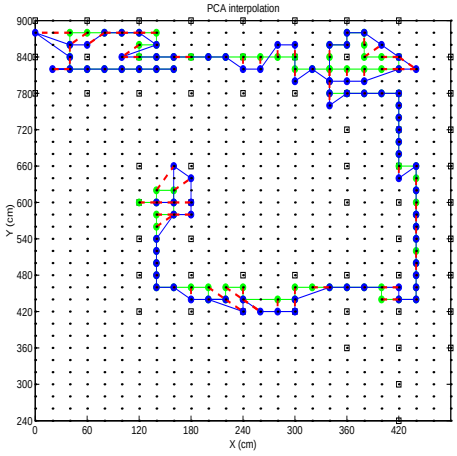


(a)

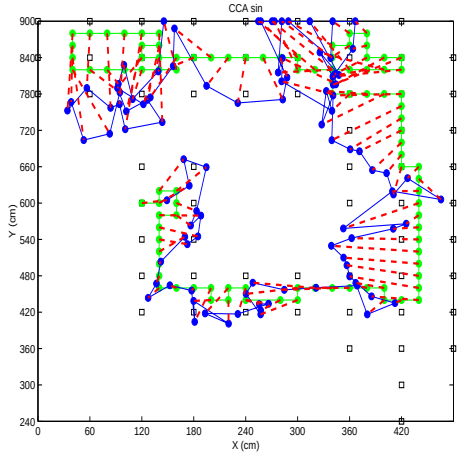
(b)



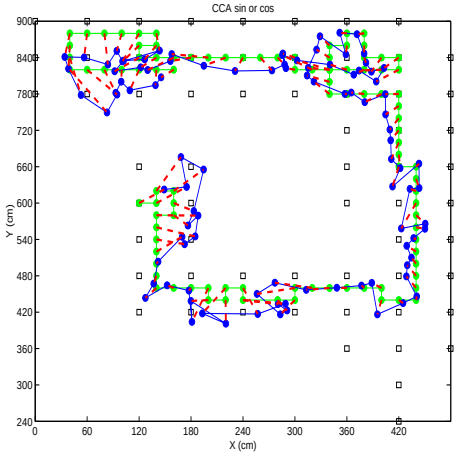
(c)



(d)



(e)



(f)

Figure 3: Localization in 2DOF: (a) CCA, (b) PCA, k=2, (c) PCA, k=10, (d) PCA, k=10, interpolated, (e) CCAsin, (f) CCAsincos.

the inverse mapping function from the feature space is unknown, we have to use a similar localization technique as in the PCA-based localization (interpolation, searching for the closest point in the subspace), thus all advantages of the CCA-based approach disappear.

A better solution would be to map the elements of location vectors using nonlinear functions whose inverse functions are known. We used sine and cosine of both offsets, and combined them together to obtain the final results. Since sine function yields better results for smaller parameter values (the errors are smaller for small values of coordinates on Fig. 3(e)), and cosine function improves the results for large parameter values, the combined results (*CCAtan*, *CCAsincos* in Fig. 3(e) and Table 1(a)) outperform the results of the standard approach. However, to further improve the results, better mapping functions will have to be used.

4.2 3DOF

In the second experiment the robot was free to move around the lab and rotate around its axis. Therefore, we had to model the (in plane) rotations also. We faced this 3DOF problem using 'spinning images' approach [3]. From each training image we generated 10 spinning images by shifting the original image to simulate the rotation. Six spanning images of one location are shown in Fig. 4. The goal was to estimate all three parameters (two offsets and rotation) using the enlarged training set.

The results are presented in Fig. 5 and Table 1(b). The CCA-based approach has proven to be useless. The PCA-based approach could not do the job using three principal vectors either, however when ten-dimensional principal subspace was used, the results were very good. It is obvious that the CCA fails to model three parameters with adequate generalization capabilities using three canonical correlation vectors only. In this case it is even more necessary to increase the number of parameters in order to make this approach useful.

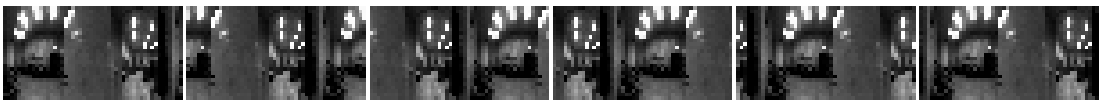
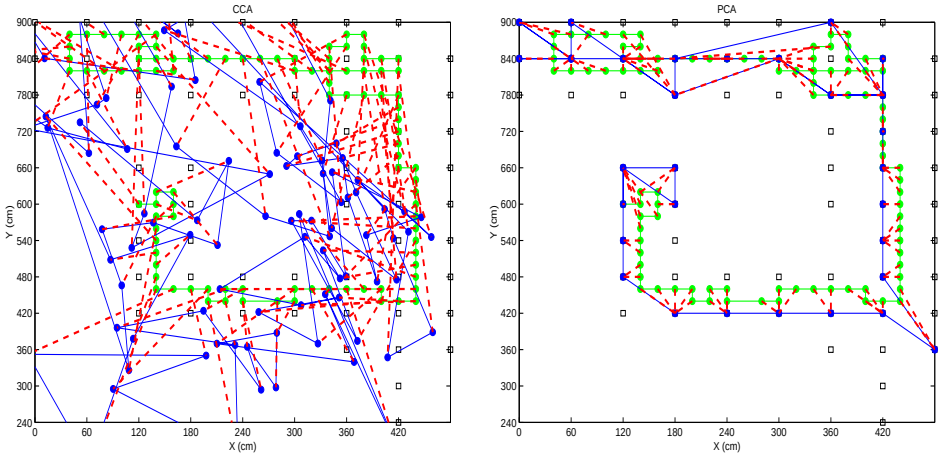


Figure 4: Six spinning images of one location.

4.3 1DOF

Then we constrained the localization problem to one degree of freedom only. The robot moved in the straight line across the lab. We captured an image every ten centimeters; 42 panoramic images of the size 58×58 altogether (see Fig. 6(a)). We used twelve of them to build the representation of the lab in the training stage and all of them in the localization stage.

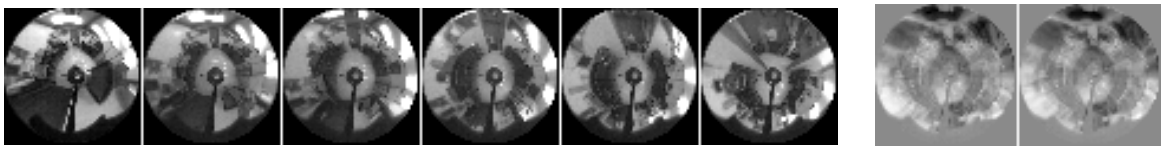
The results are shown in Fig. 7 and Table 1(c). In this case the CCA-based approach performed quite well, yielding 4.27 cm of error. However, the PCA-based approach again achieved better results when five principal vectors were used, at the cost of higher spatial and computational complexity.



(a) (b)

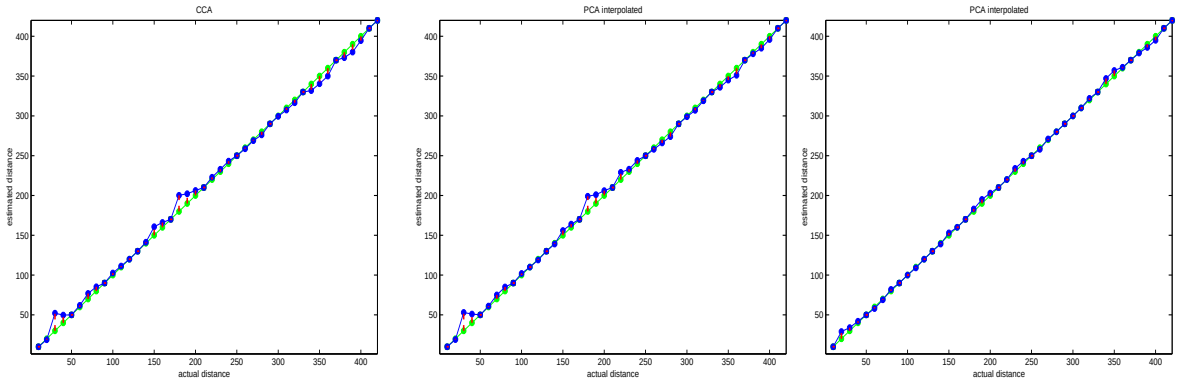
Figure 5: Localization in 3DOF: (a) CCA, (b) PCA, k=10.

It is worth noting that both approaches produce very similar subspace in this one-dimensional case. Fig. 6(b) shows the canonical correlation vector and the first principal vector, which are almost equal.



(a) (b)

Figure 6: (a) Six training images in 1DOF sequence. (b) Canonical correlation vector and first principal vector.



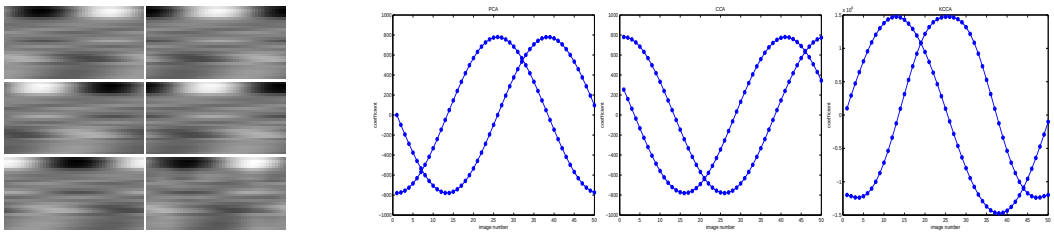
(a) (b) (c)

Figure 7: Localization in 1DOF: (a) CCA, (b) PCAint, k=1, (c) PCAint, k=5.

This is even more obvious in the following one dimensional example. In the training set

we included the spinning images, which were generated from one image only (Fig. 4). The goal was to estimate the orientation of the robot in the particular position in the room. When CCA was applied we modelled the orientation with a two-dimensional vector (sine and cosine of the angle), enabling to extract two canonical correlation vectors.

Fig. 8(a) shows the first two principal vectors, two CCA vectors and the first two KCCA vectors. One can observe, that they are very similar; the PCA and CCA vectors are equal up to the sign, while the elements in CCA vectors are only shifted. Therefore in this special case PCA and CCA produce exactly the same two-dimensional subspace. The plots in Fig. 8(b) depict PCA, CCA, and KCCA coefficients, which nicely follow sine and cosine functions. It can be shown that the principal vectors of a set of spinning images are cosine functions [7, 4]. CCA also finds the canonical correlation vectors, which map the training images into the coefficients, which are well adapted to the sine and cosine values given in the second set of observations. Such values are the most natural representations for this problem. It is also interesting that KCCA finds such representations automatically from the original one-dimensional representations of the robot orientations (rotation angles) [5].



(a) (b)

Figure 8: (a) First two PCA, CCA, and KCCA vectors, and (b) PCA, CCA, and KCCA coefficients of one set of spinning images.

(a) 2DOF	k	error	secs	elements
CCA	2	34.77	0.02	3906
PCA	2	77.37	0.21	4148
PCA	10	26.16	0.23	15044
PCAint	10	13.04	2.77	24500
KCCA	10	26.65	0.23	15044
KCCAint	10	13.12	2.81	24500
CCAsin	2	58.24	0.02	3906
CCAcos	2	34.33	0.02	3906
CCAtan	4	30.65	0.02	6520
CCAsincos	4	27.32	0.03	6520

(b) 3DOF	k	error	secs	elements
CCA	3	161.35	0.04	5212
PCA	3	210.63	2.06	8920
PCA	10	39.46	2.07	22360

(c) 1DOF	k	error	secs	elements
CCA	1	4.27	0.02	6730
PCA	1	10.00	0.05	6752
PCAint	1	3.76	0.48	7550
PCA	5	9.52	0.06	20256
PCAint	5	1.74	0.56	22650

Table 1: Results for (a) 2DOF, (b) 3DOF and (c) 1DOF.

5 Conclusion

In this paper we presented an appearance-based approach to mobile robot localization based on Canonical Correlation Analysis. The main idea is to learn the relation between the appearances of the environment at particular positions, and coordinates of these positions using CCA in the training stage, and then to use this knowledge to estimate the position of the robot in the localization stage.

The main advantages of this approach over the traditional PCA-based approach are low storage requirements and simple and very fast operation in the localization stage. However, the localization results are inferior in general. The main reason for this is that the dimensionality of the representations in CCA-based approach is limited and is very small. Thus, the representations can not contain enough information to model the environment and the generalization capabilities are not good enough. To improve the results of CCA-based localization the dimensionality of the representations should be enlarged. This would enable more information to be stored and used in the localization stage. One approach would be to map low-dimensional vectors of robot positions using non-linear functions in a high-dimensional space and perform CCA on the obtained high-dimensional data. This can be achieved using Kernel CCA, however when using kernels, the inverse mapping from the high-dimensional to the original low-dimensional space is not known. For that reason all advantages of the CCA-based approach disappear. In order to retain these advantages, the non-linear functions whose inverses are known should be used. This is the topic of our current research.

References

[1] M. Borga and H. Knutsson. Canonical correlation analysis in early vision processing. In *Proceedings of the 9th Neural Networks (ESANN)*, April 2001.

[2] G.N. de Souza and A.C. Kak. Vision for mobile robot navigation: A survey. *PAMI*, 24(2):237–267, February 2002.

[3] M. Jogan and A. Leonardis. Robust localization using an omnidirectional appearance-based subspace model of environment. *Robotics and Autonomous Sys.*, 45(1):51–72, 2003.

[4] M. Jogan, E. Žagar, and A. Leonardis. Karhunen-loeve transform of a set of rotated templates. *IEEE Trans. on Image Processing*, 12(7):817–825, 2003.

[5] T. Melzer, M. Reiter, and H. Bischof. Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 36(9):1961–1973, 2003.

[6] T. Pajdla and V. Hlavač. Zero phase representation of panoramic images for image based localization. In *CAIP'99*, pages 550–557, 1999.

[7] M. Uenohara and T. Kanade. Optimal approximation of uniformly rotated images: Relationships between Karhunen-Loeve expansion and discrete cosine transform. *IEEE Transactions on Image Processing*, 7(1):116–119, 1998.