Published in proceedings of OAGM'04, pages 15-22, 2004.

Robust Estimation of Canonical Correlation Coefficients [1)]

Danijel Skočaj *Aleš Leonardis* *Sanja Fidler*
University of Ljubljana, Slovenia
{danijel.skocaj,ales.leonardis,sanja.fidler}@fri.uni-lj.si
http://www.cogvis.fri.uni-lj.si

Abstract:
Canonical Correlation Analysis is well suited for regression tasks in appearance-based approach to modelling of objects and scenes. However, since it relies on the standard projection it is inherently non-robust. In this paper we propose to embed the estimation of CCA coefficients in an augmented PCA space, which enables detection of outliers and preserves regression-relevant information enabling robust estimation of canonical correlation coefficients.

1 Introduction

Appearance-based methods have become a popular approach to visual learning and recognition. Very often subspace methods have been used for building the representations of objects or scenes from their appearances. Among them, Canonical Correlation Analysis (CCA) [2], is best suited for regression tasks, such as estimation of objects' orientation or localization of a mobile robot.

CCA 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 an object, see Fig. 1). In the *training stage* CCA finds pairs of directions (canonical correlation vectors) that yield maximum correlation between the projections of input vectors. We can then perform linear regression on the obtained projections (canonical correlation coefficients). Later, in the *regression stage*, we can estimate the orientation (or position) of the object by using canonical correlation coefficients obtained from a novel image of the object (gray curves in Fig. 1).

[1)] 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.

Since CCA relies on the standard projection (dot product), it is inherently non-robust to non-gaussian noise. An occlusion, for instance, can cause erroneous estimation of CCA coefficients, and consequently incorrect estimation of orientation (dashed gray curve in Fig. 1). To overcome this drawback, the algorithm should be able to detect outliers and to estimate CCA coefficients from inliers only. Such approaches have already been proposed for robust estimation of PCA coefficients [5]. These algorithms take advantage of the reconstructive properties of PCA, which enable the detection of outliers. However, CCA is not a reconstructive method. 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 object) is usually low-dimensional, CCA yields only a few canonical correlation vectors, which do not enable reconstruction of input images and detection of outliers.

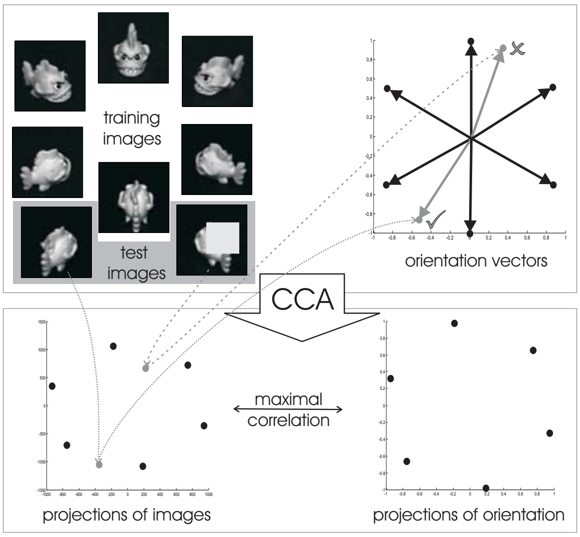


Figure 1: Principle of CCA.

Several methods for robust canonical correlation analysis have already been proposed [3], however, all of them address the robustness in the *training stage*, hence the robust estimation of canonical correlation vectors. In the training stage the training images are still available allowing to extract the information, which is necessary for outlier detection. In the *regression stage*, only the canonical correlation vectors and very low-dimensional CCA coefficients are available, thus the detection of outliers becomes practically impossible if no additional information is provided.

This is exactly what we propose in this paper. We propose to embed the estimation of CCA coefficients in the PCA space, which enables the detection of outliers. And since the truncated principal subspace may not contain all information, which is necessary for regression, we propose to augment the PCA subspace to contain the entire CCA subspace as well. Such augmented principal subspace therefore preserves information relevant for reconstruction and regression, which enables detection of outliers and reliable estimation of canonical correlation coefficients.

2 Basic CCA

We first briefly present the basic concepts of canonical correlation analysis [6] and introduce the notation.

Given N pairs of mean-normalized observations $(\hat{\mathbf{x}}_i \in \mathbb{R}^p,\ \hat{\mathbf{y}}_i \in \mathbb{R}^q)$, $i = 1, \ldots, N$, 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 $\mathbf{w}_\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{w}_\mathbf{y} \in \mathbb{R}^q$ that maximize the correlation between the projections $\mathbf{w}_\mathbf{x}^\top \hat{\mathbf{x}}_i$ and $\mathbf{w}_\mathbf{y}^\top \hat{\mathbf{y}}_i$. CCA maximizes the function

$$\rho = \frac{\mathbf{w}_\mathbf{x}^\top \mathbf{C}_\mathbf{xy} \mathbf{w}_\mathbf{y}}{\sqrt{\mathbf{w}_\mathbf{x}^\top \mathbf{C}_\mathbf{xx} \mathbf{w}_\mathbf{x} \mathbf{w}_\mathbf{y}^\top \mathbf{C}_\mathbf{yy} \mathbf{w}_\mathbf{y}}} \ , \tag{1}$$

where $\mathbf{C}_\mathbf{xx}$, $\mathbf{C}_\mathbf{yy}$, $\mathbf{C}_\mathbf{xy}$, and $\mathbf{C}_\mathbf{yx}$ are within-set and between-set covariance matrices of the input data.

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

Several approaches to maximization of (1) have been proposed. We use the dual formulation of CCA [6], which alleviates the computation in the case of high-dimensional data such as images, and solve the maximization problem with the Rayleigh quotient approach [1].

3 Robust estimation of CCA coefficients

A novel image is a subject of CCA regression, which bases its decision on the canonical correlation coefficients, i.e. projections (dot product) of the new image onto the CCA vectors. But in order to calculate the dot product properly, we need to take into account all points in an image, therefore even a small number of them being corrupted can cause erroneous results, as shown in Fig. 1.

Obviously, we would like to detect these points and exclude them from further calculations, yet still well approximate the CCA coefficients. This will be done by making use of the fact that the complete PCA basis spans *exactly* the same space as the training images, and that the canonical correlation vectors lie in this space. Therefore, we obtain *the same* results by performing CCA learning and regression on the PCA coefficients. For PCA, on the other hand, it is possible to estimate its coefficients on subsets of points. Since PCA well reconstructs the data we can efficiently pinpoint the outliers and calculate the coefficients on the rest of the image to a good degree of accuracy. It is important to note that the smaller the number

($k \ll N$) of principal vectors and coefficients is used (which still yield sufficient reconstruction of the original data), the less points in an image we need for a robust estimation. However, making the decision only on $k < N$ coefficients could in fact be disregarding discriminative information contained in the last $N - k$ of them and would lead to erroneous results. To overcome this problem we will augment the complete information necessary for CCA to the PCA truncated basis by adding to it only a small number of additional vectors.

Let $\mathbf{U_x} \in \mathbb{R}^{p \times N}$ and $\mathbf{U_y} \in \mathbb{R}^{q \times N}$ be matrices containing all N principal vectors of $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$, respectively, and let $\mathbf{A_x} \in \mathbb{R}^{N \times N}$ and $\mathbf{A_y} \in \mathbb{R}^{N \times N}$ be the matrices containing the coefficients of the training images in the PCA basis, i.e., $\mathbf{A_x} = \mathbf{U_x^\top}\hat{\mathbf{X}}$ and $\mathbf{A_y} = \mathbf{U_x^\top}\hat{\mathbf{Y}}$. By performing CCA on $\mathbf{A_x}$ and $\mathbf{A_y}$ we obtain N-dimensional canonical correlation vectors (aligned in matrices $\mathbf{V_x}$ and $\mathbf{V_y}$), which yield the same results as original CCA vectors obtained from $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$. Next, let $\mathbf{U'_x} \in \mathbb{R}^{p \times k}$, and $\mathbf{U'_y} \in \mathbb{R}^{q \times k}$ be the truncated matrices containing only the first k principal vectors. Now, the previously obtained canonical correlation vectors may not lie entirely in the subspace spanned by $\mathbf{U'_x}$. We will show how to adequately extend this subspace.

Let us first assume that the observations in the second set are only **one-dimensional**, thus $q = 1$. Therefore, CCA on $\mathbf{A_x}$ and $\mathbf{A_y}$ yields only one pair ($c = min(p, q) = 1$) of canonical correlation vectors ($\mathbf{v_x}, \mathbf{v_y}$). If we keep only k components of $\mathbf{v_x}$ (i.e., $\mathbf{v'_x} \in \mathbb{R}^{k \times 1}$), then all the information, which is contained in the last $N - k$ components of $\mathbf{v_x}$ and in the last $N - k$ principal vectors $\mathbf{u_{x}}_i$, $i = k + 1, \ldots, N$, is lost[1)]. To preserve this information we append a new basis vector $\mathbf{f}$ next to the first k principal vectors, thus obtaining a new augmented PCA basis with $k + 1$ basis vectors $\mathbf{U''_x} = \begin{bmatrix} \mathbf{U'_x} & \mathbf{f} \end{bmatrix} \in \mathbb{R}^{p \times (k+1)}$. The vector $\mathbf{f}$ is obtained in the following way:

$$\gamma = \sqrt{\sum_{i=k+1}^{N} \mathbf{v}_{\mathbf{x}i}^2} \quad , \quad \mathbf{f} = \frac{1}{\gamma} \sum_{i=k+1}^{N} v_{\mathbf{x}i}\mathbf{u}_{\mathbf{x}i} \; .$$

Since $\mathbf{f}$ is of unit length and is orthogonal to all k vectors in $\mathbf{U'_x}$, the augmented basis $\mathbf{U''_x}$ forms an orthonormal basis as well. Simultaneously we also extend $\mathbf{v'_x}$ by one element γ: $\mathbf{v''_x} = \begin{bmatrix} \mathbf{v'_x} \\ \gamma \end{bmatrix} \in \mathbb{R}^{k+1}$. Now, using $\mathbf{U''_x}$ and $\mathbf{v''_x}$, all necessary regression-relevant information is preserved.

Next we will extend this approach to the **general case** where $q > 1$. In particular, we will discuss a case which is the most common in the field of computer vision, where one set of observations contains high-dimensional images ($\mathbf{X}$) and the second set ($\mathbf{Y}$) contains low-dimensional labels or measurements of the images, thus $1 \leq q \ll N \ll p$. Therefore, CCA

1)The **boldface** characters $\mathbf{x}$ or $\mathbf{y}$ in subscript indicate that the value is related to the first ($\mathbf{X}$) or to the second set of observations ($\mathbf{Y}$), while the *italic* type denotes indices. Thus, $v_{\mathbf{x}i}$ denotes i-th element of the vector $\mathbf{v_x}$, while $\mathbf{u_y}_j$ denotes j-th vector (column) in the matrix $\mathbf{U_y}$.

yields $c = min(N, q) = q$ pairs of canonical correlation vectors $\mathbf{v}_{\mathbf{x}i}, \mathbf{v}_{\mathbf{y}i}, i = 1, \ldots, c$, aligned in the matrices $\mathbf{V}_{\mathbf{x}} \in \mathbb{R}^{N \times c}$ and $\mathbf{V}_{\mathbf{y}} \in \mathbb{R}^{q \times c}$.

Analogously to the discussion above, all information contained in the last $N - k$ principal components of all c canonical correlation vectors ($v_{\mathbf{x}ij}, i = k+1, \ldots, N,\ j = 1, \ldots, c$) has been discarded and is not contained in the truncated vectors $\mathbf{v}'_{\mathbf{x}j},\ j = 1, \ldots, c$. We will retain this information by augmenting PCA basis with additional vectors. In this case we need c additional basis vectors, which are obtained in the following way:

$$\gamma'_j = \sqrt{\sum_{i=k+1}^{N} v_{\mathbf{x}ij}^2} \quad , \quad \mathbf{f}'_j = \frac{1}{\gamma'_j} \sum_{i=k+1}^{N} v_{\mathbf{x}ij} \mathbf{u}_{\mathbf{x}i} \ ,$$

where $j = 1, \ldots, c$. Each of the vectors $\mathbf{f}'_j$ is orthogonal to all k principal vectors, however they are not mutually orthogonal. They can be orthogonalized with e.g., Gram-Schmidt orthogonalization method, i.e., $\mathbf{F} = orth(\mathbf{F}')$. Thus, after we append these additional basis vectors to the first k principal vectors ($\mathbf{U}'_{\mathbf{x}}$), we obtain an orthonormal basis of $k + c$ vectors $\mathbf{U}''_{\mathbf{x}} = \begin{bmatrix} \mathbf{U}'_{\mathbf{x}} & \mathbf{F} \end{bmatrix} \in \mathbb{R}^{p \times (k+c)}$.

Next we have to extend k-dimensional truncated canonical correlation vectors $\mathbf{v}'_{\mathbf{x}i}, i = 1, \ldots, c$ with c additional elements. In the one-dimensional case, this element (γ) simply represented the contribution of the appended basis vector $\mathbf{f}$ to the linear combination of the basis vectors. When $q > 1$ each γ'_j represents the contribution of the corresponding $\mathbf{f}'_j$. However, since we have several orthogonalized appended basis vectors, we should calculate the supplements of each truncated canonical correlation vector by considering all appended basis vectors. We extend the matrix $\mathbf{V}'_{\mathbf{x}}$ with a new matrix $\mathbf{\Gamma}$ yielding $\mathbf{V}''_{\mathbf{x}} = \begin{bmatrix} \mathbf{V}'_{\mathbf{x}} \\ \mathbf{\Gamma} \end{bmatrix} \in \mathbb{R}^{(k+c) \times c}$, where $\mathbf{\Gamma} \in \mathbb{R}^{c \times c}$ is composed of the following elements $\gamma_{ij} = \gamma'_j \langle \mathbf{f}'_j, \mathbf{f}_i \rangle = \gamma'_j \mathbf{f}_i^\top \mathbf{f}'_j$, $i = 1, \ldots, c,\ j = 1, \ldots, c$.

To summarize the complete procedure: in the *learning stage*, instead of performing CCA on the input images $\hat{\mathbf{X}}$, we can first perform PCA on these images and then CCA on the obtained coefficient vectors $\mathbf{A}_{\mathbf{x}}$. Then we may retain only the first $k \ll N$ principal vectors, providing that we append additional $c = min(p, q)$ basis vectors (obtaining $\mathbf{U}''_{\mathbf{x}}$), and adequately extend the canonical correlation vectors (obtaining $\mathbf{V}''_{\mathbf{x}}$).

In the *regression stage*, first the $(k + c)$-dimensional principal coefficient vector $\mathbf{a}_{\mathbf{x}}$ in the augmented PCA basis $\mathbf{U}''_{\mathbf{x}}$ is estimated from a novel image $\hat{\mathbf{x}}$ using the subsampling-based hypothesize-and-select robust procedure [5]. This procedure detects and discards outliers in the input image and estimates coefficients from inliers only. Then this principal coefficient vector is projected to all canonical correlation vectors (in augmented PCA basis) to obtain the canonical correlation coefficients $\mathbf{b}_{\mathbf{x}} = \mathbf{V}''_{\mathbf{x}} \mathbf{a}_{\mathbf{x}}$. Since the principal coefficients of the image are obtained in a robust way, the final canonical correlation coefficients are robust as well.

4 Experimental results

CCA is well suited for regression tasks such as estimation of objects' orientation or mobile robots' location. Due to the limited space we will present one set of experimental results for the first task only, which clearly demonstrates the advantages of our proposed method.

The experiment was performed on a set of 120 images of a toy fish, which were taken from the views evenly distributed around the object (see Fig. 2(a)). The goal was to learn the relation between the appearances of the object and their orientations using CCA in the training stage and then to use this knowledge to estimate the orientation of the object in a novel image in the test stage. Every fourth image was used in the training stage, while the remaining 90 images were used for testing.

(a) (b)

Figure 2: (a) Four non-occluded images, (b) four occluded test images.

For each training image ($\mathbf{x}_i$) its orientation (two-dimensional vector indicating the direction from which the image was taken - sine and cosine of the angle) was known ($\mathbf{y}_i$). We estimated a linear mapping from the two-dimensional vectors of canonical correlation coefficients to $\mathbf{y}_i$ using the least squares minimization method. This mapping function was then used to estimate the orientations of the test images from their canonical correlation coefficients.

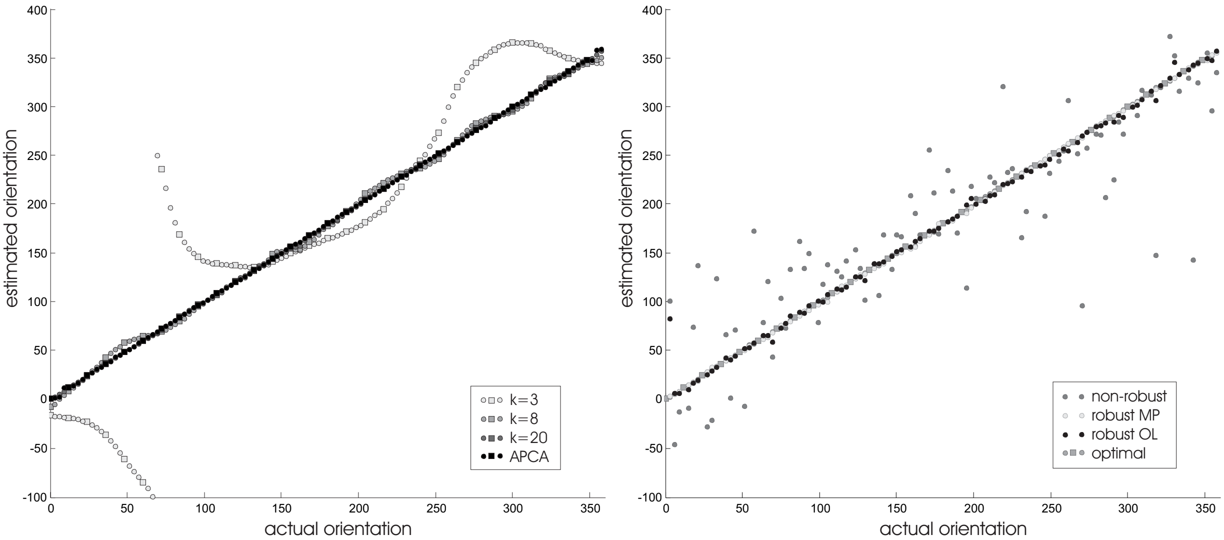
First we used non-occluded images also in the test stage. The results are shown in Fig. 3(a) and Table 1(a). The plots show the actual orientations (abscissa) and the estimated orientations (ordinate) of the object in the training images (squares) and test images (circles). The results were rather poor when we used truncated vectors of principal components (e.g., for k=3 or k=8). By increasing the number of preserved principal vectors, the results converged to the optimal result achieved either by using all principal components or augmented PCA basis or by performing CCA on the original images without the PCA preprocessing.

Then we added a square of a random intensity to a randomly chosen position in each test image (Fig. 2(b)). The results are presented in Fig. 3(b). When we used a standard non-robust method the obtained projections of occluded test images were affected by the outlying pixels, thus the estimates are very inaccurate (denoted by *non-robust*).

We then applied the PCA preprocessing step and used the robust method for estimation of coefficients. First we assumed that the positions of outliers were known; the outliers were considered as missing pixels and principal coefficients were estimated from inliers only ($robustMP$).

The results are excellent and are very close to the optimal ones.

Then we considered the test images without any additional information about outliers and applied the proposed robust method ($robustOL$). The outliers did not affect the estimation of the principal coefficients considerably, thus they did not have a significant influence on the simple projection in the regression stage either. Therefore, the results of the pose estimation are good; the deviations from the optimal estimates are mostly rather small.



(a) (b)

Figure 3: Results on (a) non-occluded and (b) occluded test images.

Table 1(b) presents mean absolute orientation errors in degrees for different dimensions of the principal subspace for the cases where the positions of outliers were assumed to be known (MP) and not known (OL). One can observe that the errors in the first case are very small even when a small number of principal components were used. As expected, the results were slightly worse when the outliers were not known, since the robust procedure first had to detect outliers. In this case a higher number of principal components should be used, since the top few principal components do not contain enough information for a reliable detection of outliers. However, these results are still significantly better than the results of the standard non-robust method.

5 Conclusion

In this paper we presented a novel approach to robust estimation of canonical correlation coefficients. We proposed to perform CCA regression in an augmented PCA space, which preserves information relevant for reconstruction (detection of outliers) and regression.

To use PCA as a preprocessing step to other subspace methods has already been proposed

Table 1: Mean absolute orientation errors: (a) non-occluded, (b) occluded test images.

k	error
3	42.08
5	5.17
8	3.61
10	2.25
15	1.25
20	0.97
APCA	0.67

(a)

k	MP	OL
3+2	2.62	20.33
5+2	2.33	14.71
8+2	1.72	6.81
10+2	1.56	4.51
15+2	1.22	3.40
20+2	1.14	2.56
non-robust		36.92

(b)

several times in the past. What is new in our approach is that we also propose to extend the truncated principal subspace with additional basis vectors. They augment the principal subspace with the information necessary for the particular subspace method (in the case of CCA this added information is a subspace spanned by canonical correlation vectors). In this way, all the properties of the subspace method are preserved and no significant information is lost.

Therefore, while the proposed method enables detection of outliers and consideration of inliers only by exploiting the reconstruction capabilities of PCA, it still preserves all necessary information. As such, the proposed approach can be used as a general tool for robustifying non-reconstructive subspace methods, such as LDA [4] or linear SVM.

References

[1] M. Borga. *Learning Multidimensional Signal Processing.* Dissertations,No. 531, Linköping University, Sweden, 1998.

[2] M. Borga and H. Knutsson. Canonical correlation analysis in early vision processing. In *ESANN'01*, April 2001.

[3] C. Dehon, P. Filzmoser, and C. Croux. Robust methods for canonical correlation analysis. *Data Analysis, Classification, and Related Methods*, pages 321–326, 2000.

[4] S. Fidler and A. Leonardis. Robust LDA classification by subsampling. In *CVPR Workshop on Statistical Analysis in Computer Vision, in press*, pages 1–8, 2003.

[5] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78:99–118, 2000.

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