ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 309-314

Canonical Correlation Analysis in Early Vision Processing

Magnus Borga Hans Knutsson

Dept. of Biomedical Engineering
Linköping University
University Hospital
SE-581 85 Linköping
Sweden

Abstract. This paper illustrates how canonical correlation analysis can be used for designing efficient visual operators by learning. The approach is highly task oriented and what constitutes the relevant information is defined by a set of examples. The examples are pairs of images displaying a strong dependence in the chosen feature but are otherwise independent. Experimental results are presented illustrating the learning of local shift invariant orientation operators, representation of velocity, and image content invariant disparity operators.

1 Introduction

The need for a generally applicable method for learning is evident in problems involving vision. The dimensionality of typical inputs often exceed 10^6, effectively ruling out any type of complete analysis. For this reason designing system capable of learning the relevant information extraction mechanisms appears to be the only possible way to proceed.

In recent years, unsupervised learning algorithms based on *mutual information* have received an increasing interest e.g. [11, 2, 3]. A set of linear basis functions, having a direct relation to maximum mutual information, can be obtained by *canonical correlation analysis* (CCA) [8]. In this paper, we present a method based on CCA for learning visual operators from examples. The following section gives a brief introduction to CCA. In section 3, the method for learning from examples is described. The method is exemplified by a number of experiments, which are presented in section 4.

2 Canonical correlation analysis

CCA finds two sets of basis vectors, one in each signal space, such that the correlation matrix between the signals described in the new basis is a diagonal matrix. A subset of the vectors containing the N first pairs defines a linear rank-N relation between the

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 309-314

sets that is optimal in a correlation sense. It has been shown that finding the canonical correlations is equivalent to maximizing the mutual information between the sets if the underlying distributions are elliptically symmetric [9].

Consider two random variables, $\mathbf{x}$ and $\mathbf{y}$, from a multi-normal distribution. Consider the linear combinations, $x = \mathbf{w}_x^T(\mathbf{x}-\bar{\mathbf{x}})$ and $y = \mathbf{w}_y^T(\mathbf{y}-\bar{\mathbf{y}})$, of the two variables respectively. $\bar{\mathbf{x}}$ denotes the mean of $\mathbf{x}$. The correlation between x and y is given by

$$\rho = \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}}. \tag{1}$$

where $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ are the nonsingular within-set covariance matrices and $\mathbf{C}_{xy}$ is the between-sets covariance matrix. The maximum of ρ with respect to $\mathbf{w}_x$ and $\mathbf{w}_y$ is the largest *canonical correlation*. A complete description of the canonical correlations is given by:

$$\begin{bmatrix} \mathbf{C}_{xx} & [\,0\,] \\ [\,0\,] & \mathbf{C}_{yy} \end{bmatrix}^{-1} \begin{bmatrix} [\,0\,] & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & [\,0\,] \end{bmatrix} \begin{pmatrix} \hat{\mathbf{w}}_x \\ \hat{\mathbf{w}}_y \end{pmatrix} = \rho \begin{pmatrix} \lambda_x \hat{\mathbf{w}}_x \\ \lambda_y \hat{\mathbf{w}}_y \end{pmatrix} \tag{2}$$

where: $\rho,\ \lambda_x,\ \lambda_y\ > 0$ and $\lambda_x\lambda_y = 1$. Equation 2 can be rewritten as:

$$\begin{cases} \mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\hat{\mathbf{w}}_y = \rho\lambda_x\hat{\mathbf{w}}_x \\ \mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\hat{\mathbf{w}}_x = \rho\lambda_y\hat{\mathbf{w}}_y \end{cases} \tag{3}$$

Solving equation 3 gives N solutions $\{\rho_n, \hat{\mathbf{w}}_{xn}, \hat{\mathbf{w}}_{yn}\},\ n = \{1..N\}$. N is the minimum of the input dimensionality and the output dimensionality. ρ_n are the *canonical correlations* [8]. More details can be found in [4].

3 Learning visual operators from examples

The basic idea behind the proposed method, illustrated in figure 1, is to analyse two signals where the feature that is to be represented generates dependent signal components. The signal vectors fed into the CCA are image data mapped through a function f. In general, f can be any vector-valued function of the image data. The choice of f is of major importance as it determines the representation of input data for the canonical correlation analysis. It is f that gives the desired invariance properties. Other authors have proposed nonlinear extensions to CCA, which includes f in the learning process [1, 10]. In this case, however, we have used a fixed function f.

The training data are presented in pairs such that the features for which we want to find a representation vary in a correlated way. Other features, for which we want the representation to be invariant to, are varied in an unordered way. In this way, the desired features are captured by the CCA.

4 Experimental results

The proposed method is exemplified by a number of experiments. Due to the limited space, the descriptions of the experiments are very short. For more information, please see the references.

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 309-314

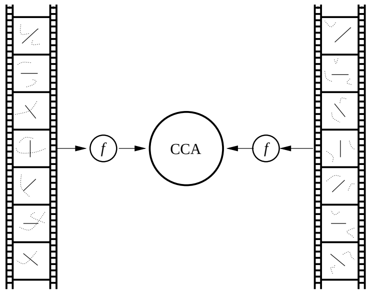

Figure 1: A symbolic illustration of the method of using CCA for finding visual operators.

4.1 Local orientation

It is shown in [4] and [7] that if f is an outer product and the image pairs contain sine wave patterns with equal orientations but different phase, the CCA finds linear combinations of the outer products that convey information about local orientation and are invariant to local phase. Figures 2 and 3 show results from a similar experiment, this time using image pairs of edges having equal orientation and different, independent positions. Independent white Gaussian noise to a level of 12 dB SNR was added to all images before training. The noise forces the CCA to develop robust operators with as little noise-sensitivity as possible.

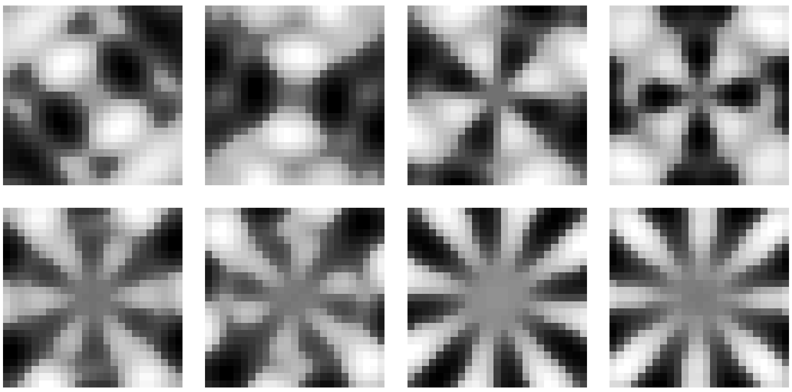

Figure 2: Projections of Fourier components on canonical correlation vectors 1 to 8.

Figure 2 show the projections of Fourier components on canonical correlation vectors 1 to 8. The result shows that angular operators of orders 2, 4, 6 and 8 have been formed and are important information carriers. The magnitude of the projections are close to shift-invariant having a position dependent variation in the order of 5 %.

Performing an eigenvalue decomposition of the canonical correlation vectors the

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 309-314

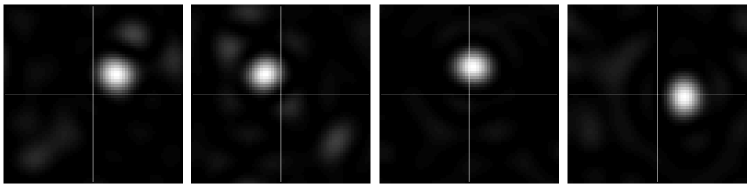

Figure 3: Spectra of eigenimages interpreted as complex quadrature filter pairs.

corresponding linear combinations, in the outer product space, can be seen as quadratic combinations of linear filters [4]. The linear filters (eigenimages) obtained display a clear tendency to form pairs of odd and even filters having similar spectra. Such quadrature filter pairs allow for a local shift-invariant feature and are functionally similar to the orientation selective ‘complex cells’ found in biological vision. Figure 3 shows the spectra of four such filter pairs. The two left spectra are from canonical correlation vector one and display selectivity to orientations 45 and 135 degrees. The two right spectra are from canonical correlation vector two and display selectivity to orientations 0 and 90 degrees.

4.2 Generation of motion representation

The basic idea in this experiment is to feed filter output to the CCA in such a way that velocity is the only common information in the two inputs. Therefore the input is taken from two windows moving with the same velocity and in the same direction. Each window consists of 9 points i a square grid.

The steps were taken with a velocity vector that was randomly changed every step with a change of -1, 0 or 1 in each direction. The norm of the velocity vector was limited to 5 (pixels per step). For each step, quadrature filter output from the image was multiplied by the conjugate of the corresponding filter output from the previous step. The multiplication with the conjugate extracts the phase difference between the two steps which contains information about the motion. 8 filters were used which gives 72-dimensional complex vectors. These vectors were then normalized and divided into a real and an imaginary part. The resulting pair of a 144-dimensional vector from each window was then used as input to the CCA.

To visualise the resulting velocity representation, the filter response vectors for different velocities were projected onto the three most significant CCA-vectors. In figure 4 (left), the projections onto the first two CCA-vectors are plotted and at each point, the true direction of motion is indicated with an arrow. The clusters in the middle of the figure are caused by the quantization of the velocity in whole pixels per time unit. Figure 4 (middle) shows the projections onto the first and the third CCA-vector. In this plot, also the magnitude of the true velocity is indicated as the length of the arrow. From these two views, we see that the CCA has generated a three-dimensional representation of the velocity approximately on the surface of a sphere. In the first two dimensions, the *direction* is represented, while the *magnitude* of the velocity also uses the third dimension. A stylised version of the generated 3D

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 309-314

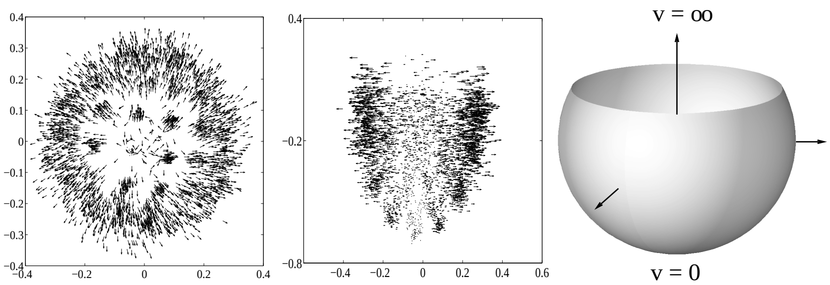

Figure 4: Projections of motion induced signals onto the first two canonical vectors (left) and the first and the third canonical vectors (middle). The true direction of motion is indicated as arrows. To the right is a stylized version of the generated 3D velocity representation.

Figure 5: Result on a stereo pair of Pentagon.

representation is shown to the right in figure 4. For more information, see [5].

4.3 Local disparity

An important problem in computer vision that is suitable to handle with CCA is stereo vision, since data in this case naturally appear in pairs. In [4, 6], a novel stereo vision algorithm that combines CCA and phase analysis is presented. It is demonstrated that the algorithm can handle traditionally difficult problems such as: **1.** Producing multiple disparity estimates for semi-transparent images, **2.** Maintain accuracy at disparity edges, and **3.** Allowing differently scaled images.

Canonical correlation analysis is used to create adaptive linear combinations of quadrature filters. These linear combinations are new quadrature filters that are adapted in frequency response and spatial position so that the correlation between the filter outputs from the two images is maximized. Figure 5 shows the resulting disparity estimates on a well known test image pair.

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 309-314

5 Conclusions

We have presented a method for learning visual operator by canonical correlation analysis. Experimental results indicate that the proposed method is applicable on a wide range of different visual tasks.

References

[1] H. Asoh and O. Takechi. An approximation of nonlinear canonical correaltion analysis by multilayer perceptrons. In *Proceedings of the International Conference on Artificial Neural Networks*, pages 713–716, Sorrento, 1994.

[2] S. Becker and G. E. Hinton. Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355(9):161–163, January 1992.

[3] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–59, 1995.

[4] M. Borga. *Learning Multidimensional Signal Processing*. PhD thesis, Linköping University, Sweden, SE-581 83 Linköping, Sweden, 1998. Dissertation No 531, ISBN 91-7219-202-X.

[5] M. Borga, M. Andersson, and H. Knutsson. Generation of representations for supervised learning - a velocity estimation example. In *Proceedings of the 12th Scandinavian Conference on Image Analysis*, Bergen, Norway, June 2001. Submitted.

[6] M. Borga and H. Knutsson. Estimating multiple depths in semi-transparent stereo images. In *Proceedings of the Scandinavian Conference on Image analysis*, Greenland, June 1999. SCIA. Also as Technical Report LiTH-ISY-R-2248.

[7] M. Borga, H. Knutsson, and T. Landelius. Learning canonical correlations. In *Proceedings of the 10th Scandinavian Conference on Image Analysis*, Lappeenranta, Finland, June 1997. SCIA.

[8] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[9] J. Kay. Feature discovery under contextual supervision using mutual information. In *International Joint Conference on Neural Networks*, volume 4, pages 79–84. IEEE, 1992.

[10] P. L. Lai and C. Fyfe. A neural implementatoion of canonical correlation analysis. *Neural Networks*, 12:1391–1397, 1999.

[11] R. Linsker. Development of feature-analyzing cells and their columnar organization in a layered self-adaptive network. In Rodney M. L. Cotteril, editor, *Computer Simulation in Brain Science*, chapter 27, pages 416–431. Cambridge University Press, 1988.