

Finding Efficient Nonlinear Visual Operators using Canonical Correlation Analysis

Magnus Borga Hans Knutsson
magnus@isy.liu.se *knutte@isy.liu.se*

Computer Vision Laboratory
Dept. of Electrical Engineering
Linköping University
SE-585 94 Linköping
Sweden

Abstract

This paper presents a general strategy for designing efficient visual operators. 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. Particularly important concepts in the work are mutual information and canonical correlation. Visual operators learned from examples are presented, e.g. local shift invariant orientation operators and image content invariant disparity operators. Interesting similarities to biological vision functions are observed.

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. In common practice vision problems are handled by reducing the dimensionality to typically < 10 by throwing away almost all available information in a basically *ad hoc* manner. This approach is however likely to fail if, as is frequently the case for vision problems, the mechanisms by which the necessary information can be extracted is not well understood. For this reason designing system capable of learning the relevant information extraction mechanisms appears to be the only possible way to proceed.

Hebbian learning methods like Oja's rule [12] and self-organizing feature maps [10] are related to the principal components of the signal distribution and, hence, base their selection of basis vectors on signal variance.

However, when the problem involves an analysis of the relations between two sets of data, the principal components of either set are not relevant. In recent years, unsupervised learning algorithms based on *mutual information* [13] have received an increas-

ing interest. Examples of this approach are the *informax principle* [11], *Imax principle* [2] and Independent Components Analysis [7]. Mutual information based learning has been used e.g. for blind separation and blind deconvolution [3] and disparity estimations in random-dot stereograms [2].

A set of linear basis functions, having a direct relation to maximum mutual information, can be obtained by *canonical correlation analysis* (CCA) [8]. CCA finds two sets of basis functions, one in each signal space, such that the correlation matrix between the signals described in the new basis is a diagonal matrix. The basis vectors can be ordered such that the first pair of vectors \mathbf{w}_{x1} and \mathbf{w}_{y1} maximizes the correlation between the projections ($\mathbf{x}^T \mathbf{w}_{x1}$, $\mathbf{y}^T \mathbf{w}_{y1}$) of the signals \mathbf{x} and \mathbf{y} onto the two vectors respectively. A subset of the vectors containing the N first pairs defines a linear rank- N relation between the sets that is optimal in a correlation sense. In other words, it gives the linear combination of one set of variables that is the best predictor and at the same time the linear combination of an other set which is the most predictable. 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].

2 Canonical correlation analysis

Consider two random variables, \mathbf{x} and \mathbf{y} , from a multi-normal distribution:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N \left(\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix} \right), \quad (1)$$

where $\mathbf{C} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix}$ is the covariance matrix. \mathbf{C}_{xx} and \mathbf{C}_{yy} are nonsingular matrices and $\mathbf{C}_{xy} = \mathbf{C}_{yx}^T$. Consider the linear combinations, $x = \mathbf{w}_x^T (\mathbf{x} - \mathbf{x}_0)$ and $y = \mathbf{w}_y^T (\mathbf{y} - \mathbf{y}_0)$, of the two variables respectively. The correlation between x and y is given by

(2), see for example [1]:

$$\rho = \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}}. \quad (2)$$

A complete description of the canonical correlations is given by:

$$\begin{bmatrix} \mathbf{C}_{xx} & [0] \\ [0] & \mathbf{C}_{yy} \end{bmatrix}^{-1} \begin{bmatrix} [0] & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & [0] \end{bmatrix} \begin{pmatrix} \hat{\mathbf{w}}_x \\ \hat{\mathbf{w}}_y \end{pmatrix} = \rho \begin{pmatrix} \lambda_x \hat{\mathbf{w}}_x \\ \lambda_y \hat{\mathbf{w}}_y \end{pmatrix} \quad (3)$$

where: $\rho, \lambda_x, \lambda_y > 0$ and $\lambda_x \lambda_y = 1$. Equation (3) can be rewritten as:

$$\begin{cases} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \hat{\mathbf{w}}_y = \rho \lambda_x \hat{\mathbf{w}}_x \\ \mathbf{C}_{yy}^{-1} \mathbf{C}_{yx} \hat{\mathbf{w}}_x = \rho \lambda_y \hat{\mathbf{w}}_y \end{cases} \quad (4)$$

Solving (4) 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. The linear combinations, $x_n = \hat{\mathbf{w}}_{xn}^T \mathbf{x}$ and $y_n = \hat{\mathbf{w}}_{yn}^T \mathbf{y}$, are termed *canonical variates* and the correlations, ρ_n , between these variates are termed the *canonical correlations* [8]. An important aspect in this context is that the canonical correlations are *invariant*

to affine transformations of \mathbf{x} and \mathbf{y} . Also note that the canonical variates corresponding to the different roots of (4) are uncorrelated, implying that:

$$\begin{cases} \mathbf{w}_{xn}^T \mathbf{C}_{xx} \mathbf{w}_{xm} = 0 \\ \mathbf{w}_{yn}^T \mathbf{C}_{yy} \mathbf{w}_{ym} = 0 \end{cases} \quad \text{if } n \neq m \quad (5)$$

It should be noted that (3) is a special case of the *generalized eigenproblem* [4]:

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

3 Learning from Examples

Descriptors for higher order features are in practice impossible to design by hand due to the overpowering amount of possible signal combinations. In [4] it is shown how canonical correlation analysis can be used to find operators that represent relevant local features in images.

The basic idea behind the CCA approach, 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 . If f is the identity operator (or any other full-rank linear function), the CCA finds the linear combinations of pixel data that have the highest correlation. In this case, the canonical correlation vectors can be seen as linear filters. In general, f can be any vector-valued function of the image data, or even different functions f_x and f_y , one for each signal space. The choice of f is of major importance as it determines the representation of input data for the canonical correlation analysis.

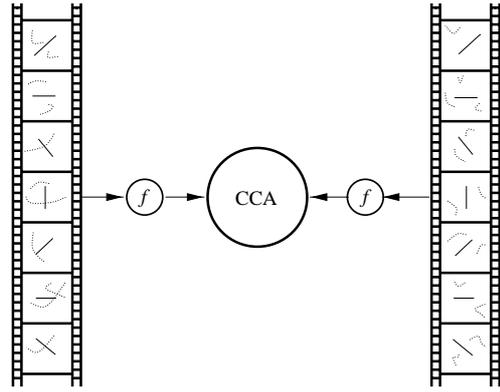


Figure 1: A symbolic illustration of the method of using CCA for finding feature detectors in images. The desired feature (orientation: here illustrated by a solid line) is varying equally in both image sequences while other features (here illustrated with dotted curves) vary in an uncorrelated way. The input to the CCA is a function f of the image.

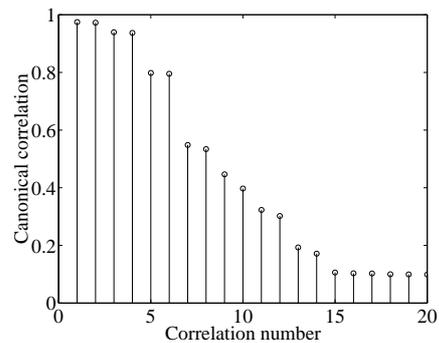


Figure 2: Plot of canonical correlation values.

Local orientation

It is shown in [4] and [6] 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, 3 and 4 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. Figure 2 shows the values of the 20 first canonical correlations. The values appear to come in pairs, the first two values being ≈ 0.98 demonstrating that the mutual information mediated through local orientation is high.

Figure 3 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 has been formed and are important information carriers. The magnitude of the projections are close

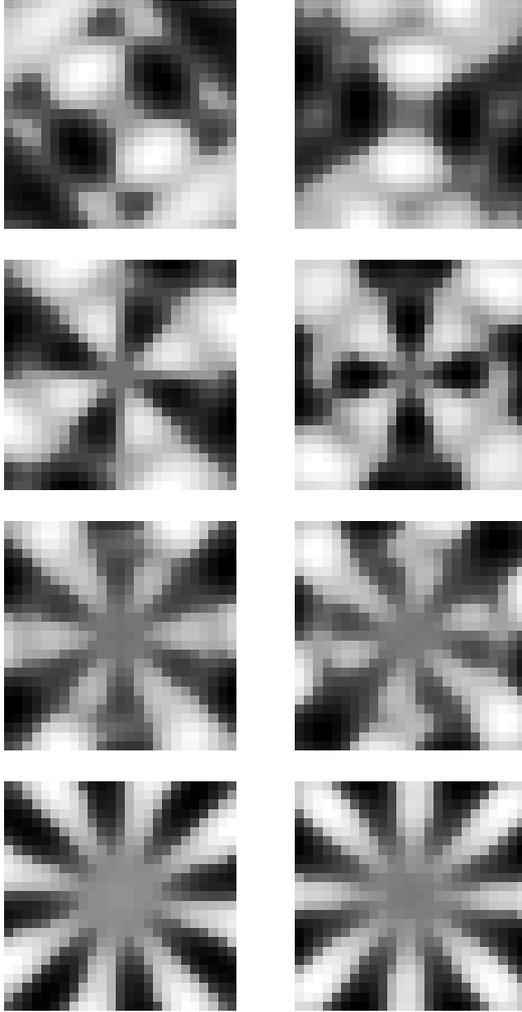


Figure 3: Projections of Fourier components on canonical correlation vectors 1 to 8. The result shows that angular operators of orders 2, 4, 6 and 8 are important information carriers.

to shift-invariant having a position dependent variation in the order of 5 %. Comparing to figure 2 it can be seen that the decrease in the canonical correlation values corresponds to an increase in angular order of the operators.

‘Complex cells’ Performing an eigenvalue decomposition of the canonical correlation vectors the 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 4 shows the spectra of four such filter pairs. The top two are from canonical correlation vector one and display selectivity to orientations 45 and 135 deg. The bottom two are from canonical correlation vector two and display selectivity to orientations 0 and 90 deg.

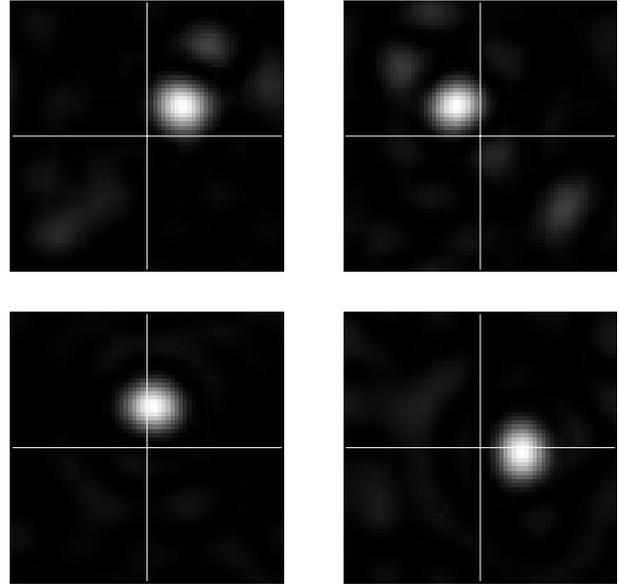


Figure 4: Spectra of eigenimages interpreted as complex quadrature filter pairs. Top two from canonical correlation vector 1. Bottom two from canonical correlation vector 2.

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, 5] 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, see figure 6, **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 maximizing the correlation between the filter outputs from the two images. Figure 5 shows the filters obtained for two white noise images where disparity changes linearly with horizontal position. Note that the obtained filters have adapted to the effect of the disparity gradient through a relative offset in center frequency. The disparity estimate is obtained by analysing the phase of the scalar product of the adapted filters. A result for depth estimates using semi-transparent images is shown in figure 6.

4 The Future

The concept of mutual information provides a solid and general basis for the study of a broad spectrum of problems including signal operator design and learning strategies.

A broadly applicable general approach, illustrated

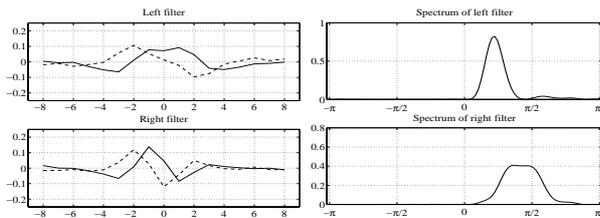


Figure 5: The filters created by CCA. Solid lines show the real parts and dashed lines show the imaginary parts.

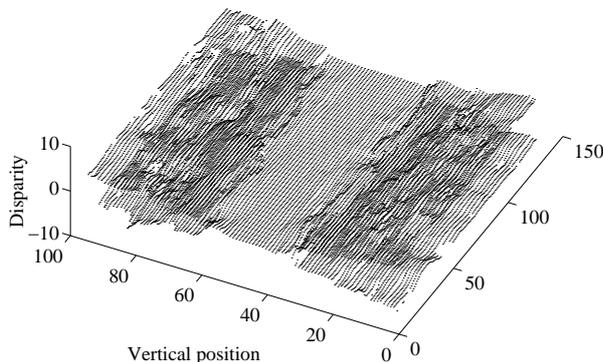


Figure 6: The result of the stereo algorithm for two random dot images corresponding to two semi-transparent crossing planes.

in figure 7, is to maximize mutual information subject to constraints given by a chosen model space. This could be done by varying not only the linear projections, i.e. the CCA part, but also the functions f_x and f_y .

Finding suitable function classes and efficient parameterisation/implementations for these functions is still the central issue and will be an important theme in our continued investigations.

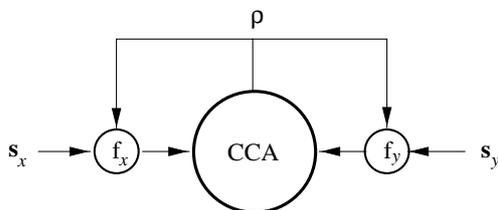


Figure 7: A general approach for finding maximum mutual information.

Acknowledgement

We like to thank The Swedish Research Council for Engineering Sciences (TFR) and The Swedish National Board for Industrial and Technical Development.

References

- [1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, second edition, 1984.
- [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 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.
- [6] 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.
- [7] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, April 1994.
- [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] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.
- [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.
- [12] E. Oja. A simplified neuron model as a principal component analyzer. *J. Math. Biology*, 15:267–273, 1982.
- [13] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 1948. Also in N. J. A. Sloane and A. D. Wyner (ed.) *Claude Elwood Shannon Collected Papers*, IEEE Press 1993.