An Adaptive Stereo Algorithm Based on Canonical Correlation Analysis

M. Borga, H. Knutsson
Computer Vision Laboratory
Department of Electrical Engineering
Linköping University
S-581 83 Linköping, Sweden

Abstract – **This paper presents a novel algorithm that uses CCA and phase analysis to detect the disparity in stereo images. The algorithm adapts filters in each local neighbourhood of the image in a way which maximizes the correlation between the filtered images. The adapted filters are then analysed to find the disparity. This is done by a simple phase analysis of the scalar product of the filters. The algorithm can even handle cases where the images have different scales. The algorithm can also handle depth discontinuities and give multiple depth estimates for semi-transparent images.**

I. Introduction

An important problem in most signal processing tasks is to find the best basis to represent the signal. Many unsupervised learning methods have been proposed to handle this problem. Hebbian learning methods like Oja's rule [15] and self-organizing feature maps [12] 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* [17] have received an increasing interest. Examples of this approach are the *infomax principle* [13] and the *Imax principle* [3]. Mutual information based learning has been used for blind separation and blind deconvolution [5] and disparity estimations in random-dot stereograms [3].

A set of linear basis functions, having a direct relation to maximum mutual information, can be obtained by *canonical correlation analysis* (CCA) [9]. 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 [11].

An iterative algorithm for successively finding the canonical correlations and the corresponding vectors has been presented [7], [8]. It has been shown that CCA can be used to find image feature detectors that describe a particular feature while being invariant to other features [7], [8]. The features to be described are learned by the algorithm by giving examples that are presented in pairs in such a way that the desired feature, e.g. orientation, is equal for each pair while other features, e.g. phase, are presented in an unordered way. It has also been shown that CCA can be used to find higher order (quadratic) combinations of filter outputs from a local neighbourhood to get orientation estimates that are less sensitive to noise than the standard vector averaging method [7].

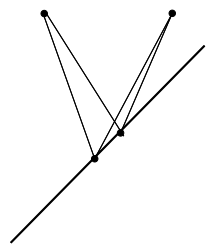

Fig. 1. Scaling effect when viewing a tilted plane.

An important problem in computer vision that is suitable to handle with this technique is stereo vision, since data in this case naturally appears in pairs. The problem of estimating disparity between pairs of stereo images, is not a new problem [2]. Earlier approaches often used matching of some feature in the two images [14]. Later approaches have been more focused on using the phase information given by for example Gabor or quadrature filters [16], [19], [10], [18]. An advantage with phase based methods is that phase is a continuous variable that allows for sub-pixel accuracy.

A problem with phase based stereo algorithms is that phase estimation works only within a relatively small frequency band. To solve this problem, the disparity estimation can be made sequentially beginning on a coarse scale and gradually refining the estimates on finer scales [18]. A problem that is not solved by this approach is when the observed surface is tilted in depth so that the depth varies along the horizontal axis. In this situation, the surface will be viewed at different scales by the two cameras as illustrated in Fig. 1. This means that phase information on one scale in the left image must be compared with phase information on another scale in the right image. In most stereo algorithms this problem cannot be handled in a simple way.

Another problem that most stereo algorithms are faced with occurs at vertical depth discontinuities (but see [4]). Around the discontinuity there will be a region where the algorithm either will not be able to make any estimate at all, or the estimate will be some average between the two correct disparities which indicates a slope rather than a step.

This paper presents a stereo algorithm that solves these two important problems by a combination of phase and canonical correlation analysis. It can also give multiple estimates for semi-transparent images. Examples of such images are X-ray images. The algorithm can, in a simplified way, be described as follows: 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 in order to maximize the correlation (i.e. linear agreement) between the filter outputs in the two images. The coefficients given by the canonical correlation vectors are then used as weighting coefficients in a pre-computed table that allows for an efficient phase-based search for the disparity (or disparities).

The following section gives a brief review of the theory of canonical correlation. In section III, the concept of quadrature filters and phase is described. In section IV the stereo algorithm is presented and, finally, in section V some experimental results are shown.

II. Canonical correlation analysis

Consider two random variables, $\mathbf{x}$ and $\mathbf{y}$, from a multinormal distribution:

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

where $\mathbf{C} = \left[\begin{smallmatrix}\mathbf{C}_{xx} & \mathbf{C}_{xy}\\ \mathbf{C}_{yx} & \mathbf{C}_{yy}\end{smallmatrix}\right]$ 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]:

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

A complete description of the canonical correlations is given by:

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

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

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

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* [9]. 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:

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

It should be noted that (3) is a special case of the *generalized eigenproblem* [7]:

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

III. Quadrature and phase

There are many compelling arguments for designing edge and line detection filters as pairs with one even part (line detector) and one odd part (edge detector). The even part can roughly be thought of as a period of a cosine wave centred around zero and the odd part as a period of a sine wave also centred around zero. A convenient way to handle such filter pairs is to combine them into complex filters, where the real part is the even filter and the imaginary part is the odd filter. The type of line/edge event can then be represented as the phase of the complex filter output: zero phase indicates that the filter is on the middle of a bright line, 90° indicates the middle of an edge and so on.

If the magnitude of the filter output is invariant with respect to phase when applied to a pure sine wave pattern, the filter is defined as a *quadrature filter*. The Fourier transform of such a filter is zero in one half plane (for example negative frequencies in the one-dimensional case) and have zero DC component. This means that any linear combination of quadrature filters in the same direction is also a quadrature filter since the negative frequency response and DC component remain zero.

IV. The stereo algorithm

The algorithm consists of two parts; CCA and phase analysis. Both analyses are performed for each disparity estimate.

A. Canonical correlation analysis part

The input $\mathbf{x}$ and $\mathbf{y}$ to the CCA come from the left and right images respectively. Each input is a vector with outputs from a set of quadrature filters. In the implementation described here, this set consists of two identical one-dimensional (horizontal) quadrature filters with two pixels relative displacement. (Of course other and larger sets of filters can be used including, for example, filters with different bandwidths, different centre frequencies, etc.)

The data is sampled from a neighbourhood $\mathcal{N}$ around the point for the disparity estimate. The choice of neighbourhood size is a compromise between noise sensitivity and locality. The covariance matrix $\mathbf{C}$ is calculated using the vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathcal{N}$. The fact that quadrature filters have zero DC component simplifies this calculation to an outer product sum:

$$\mathbf{C} = \sum_{\mathcal{N}} \begin{pmatrix}\mathbf{x}_i\\ \mathbf{y}_i\end{pmatrix} \begin{pmatrix}\mathbf{x}_i\\ \mathbf{y}_i\end{pmatrix}^T \tag{6}$$

The first canonical correlation ρ_1 and the corresponding vectors $\mathbf{w}_x$ and $\mathbf{w}_y$ are then calculated by solving (4). In the case where only two filters are used, this calculation becomes very simple. If very large sets of filters are used, the covariance matrix gets very big and an analytical calculation of the canonical correlation becomes computationally very expensive. In such a case, an iterative $O(n)$ algorithm, that avoids outer products and matrix inverses, can be used [7], [8].

The canonical correlation vectors define two new filters, $\mathbf{f}_x = \sum_{i=1}^{N} w_{xi}\mathbf{f}_i$ and $\mathbf{f}_y = \sum_{i=1}^{N} w_{yi}\mathbf{f}_i$ where $\mathbf{f}_i$ are the basis filters, N is the number of filters in the filter set and w_{xi} and w_{yi} are the components in the first pair of canonical correlation vectors. $\mathbf{f}_x$ and $\mathbf{f}_y$ have maximum correlated output in $\mathcal{N}$, given the set of basis filters $\mathbf{f}_i$.

B. Phase analysis part

The key idea behind this part is to search for the disparity that corresponds to a real valued correlation between the two new filters. This is based on the fact that canonical correlations are real valued [7]. In other words, we want to find the disparity δ such that the correlation coefficient function [6] between the new filters' output, q_x and q_y, is real valued, i.e.

$$\mathrm{Im}\left[\mathrm{Corr}\left(\mathrm{q_x}(\xi)\ ,\ \mathrm{q_y}(\xi+\delta)\right)\right] = 0 \tag{7}$$

where ξ is the spatial (horizontal) coordinate.

A calculation of the correlation over $\mathcal{N}$ for all δ would be very expensive. A much more efficient solution is to make the simplified assumption that the signal $\mathbf{s}$ can be modelled by the covariance matrix $\mathbf{C}_{ss}$. This assumption reduces the correlation calculation to a scalar product between the filters coefficient vectors weighted by the signal covariance matrix. The goal then is to find the solution to

$$\mathrm{Im}\left[\frac{\mathbf{f}_x^*\mathbf{C}_{ss}\mathbf{f}_y(\delta)}{\sqrt{\mathbf{f}_x^*\mathbf{C}_{ss}\mathbf{f}_x\ \mathbf{f}_y^*\mathbf{C}_{ss}\mathbf{f}_y}}\right] = 0 \tag{8}$$

where $(^*)$ denotes conjugate transpose and $\mathbf{f}_y(\delta)$ is a shifted version of $\mathbf{f}_y$.

A lot of the computations needed to solve (8) can be saved since

$$\begin{aligned}\mathbf{f}_x^*\mathbf{C}_{ss}\mathbf{f}_y(\delta) &= \left(\sum_{i=1}^{N} w_{xi}\mathbf{f}_i\right)^* \mathbf{C}_{ss}\left(\sum_{j=1}^{N} w_{yj}\mathbf{f}_j(\delta)\right)\\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} w_{xi}^* w_{yj}\mathbf{f}_i^*\mathbf{C}_{ss}\mathbf{f}_j(\delta) = \sum_{ij} v_{ij} g_{ij}(\delta)\end{aligned} \tag{9}$$

where $g_{ij}(\delta) = \mathbf{f}_i^*\mathbf{C}_{ss}\mathbf{f}_j(\delta)$ can be calculated in advance for different disparities δ and stored in a table. In the case of two basis filters the table contains four rows. Hence, for a given disparity a (complex) correlation can be computed as a normalized sum of the four values from the table weighted with the coefficients v_{ij} given by the CCA.

The aim is now to find the δ's for which this correlation is real valued. This is done by finding the zero crossings of the phase of the correlation. This can be done using a very coarse quantization of δ in the table since the phase is, in general, rather linear near the zero crossing (as opposed to the imaginary part which in general is not linear). Hence, first coarse estimates of the zero crossings are obtained. Then the derivatives of the phases at the zero crossings are measured. Finally, the errors in the coarse estimates are compensated for by using the actual phase values and the phase derivatives at the estimated positions:

$$\delta = \delta_c - \frac{\varphi(\delta_c)}{\partial\varphi/\partial\delta} \tag{10}$$

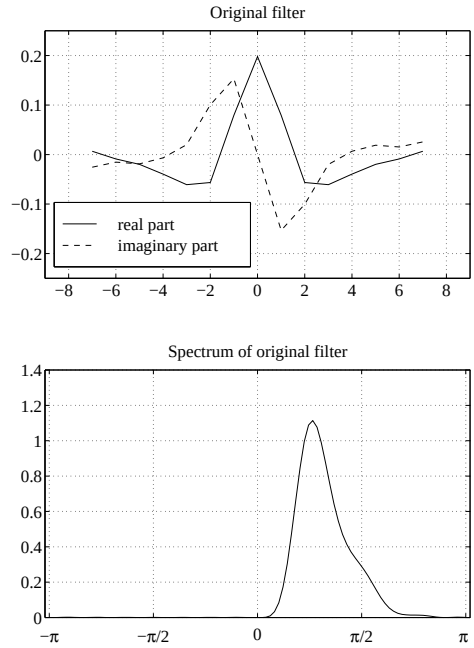

Fig. 2. The filter in the basis filter set. Solid lines show the real parts and dashed lines show the imaginary parts.

where δ_c is the coarse estimate of the zero crossing and $\varphi(\delta_c)$ is the phase at that position.

If more than one zero crossing is detected, the magnitudes of the correlations can be used to select a solution. Since the CCA searches for maximum correlation, the zero crossing with maximum correlation is most likely to be the best estimate. If two zero crossings have approximately equal magnitudes, both disparity estimates can be considered to be correct within the neighbourhood. This indicates a depth discontinuity.

If the images are differently scaled, the CCA tries to create filters scaled correspondingly to fit the signals. In order to improve the disparity estimates in these cases, the table can be extended with scaled versions of the basis filters. The CCA step is not affected by this and the phase analysis is performed as described above on each scale. The correct scale is indicated by having the maximum real valued correlation. The resolution in scale can be very coarse. In our experiments, we have used filters scaled between $+/-$ one octave in steps of a quarter of an octave, which seems to be quite a sufficient resolution.

V. Experimental results

In all experiments presented here a basis filter set have been used consisting of two one-dimensional horizontally oriented quadrature filters, both with a centre frequency of $\pi/4$ and a bandwidth of two octaves. The filters have 15 coefficients in the spatial domain and are shifted two pixels relative to each other. The frequency function is a quadratic cosine on a log scale:

$$F(u) = \cos^2\left(k\ln(u/u_0)\right) \tag{11}$$

where $k = \pi/\left(2\ln(2)\right)$ and $u_0 = \pi/4$. The filter is illustrated in Fig. 2.

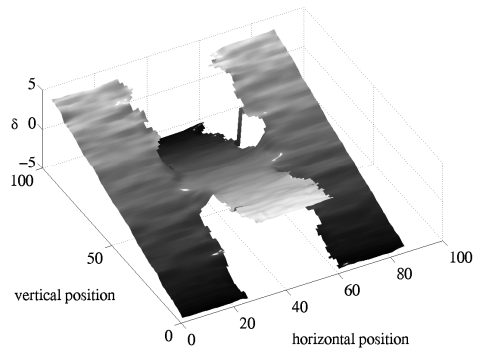

Fig. 3. Disparity estimate for different depth discontinuities.

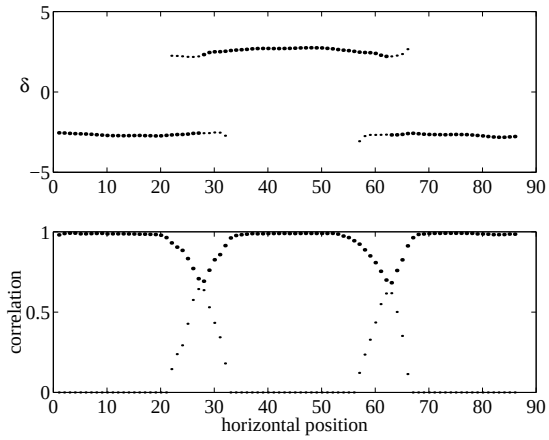

Fig. 4. Top: line 20 from the disparity estimates in Fig. 3. The small dots indicates the disparity estimates with the second strongest correlations. Bottom: The corresponding correlations.

The first experiment illustrates the algorithm's ability to handle depth discontinuities. The test image is made of white noise shifted so that the disparity varies between $+/-d$ along the horizontal axis and d varies as a ramp from -5 pixels to $+5$ pixels along the vertical axis in order to get discontinuities between $+/-10$ pixels. A neighbourhood $\mathcal{N}$ of 13×7 pixels was used for the CCA. Fig. 3 shows the estimated disparity for this test image. Disparity estimates with a correlation less than 0.7 at the zero crossing have been removed. In Fig. 4 and Fig. 5, two lines from the disparity estimate are shown. In Fig. 4, line 20 with a disparity of $+/-2.5$ pixels is shown and in Fig. 5¡, line 38 with a disparity of 1 pixel is shown. The figures at the top shows the most likely (large dots) and second most likely (mall dots) disparity estimates along these lines. The bottom figures show the corresponding correlations at the zero crossings. Fig. 3, Fig. 4 and Fig. 5 shows that for small discontinuities, the algorithm interpolates the estimates while for large discontinuities, there are two overlapping estimates.

The second experiment shows that the algorithm can estimate disparities between images that are differently scaled. The test image here is white noise warped to form a ramp along the horizontal axis. The warping is made so that the right image is compressed 50% which means that there is a scale difference of one octave. For a human, this corresponds to looking at a point on a surface with its normal rotated 67° from the observer at a distance of 20 centimetres. Here, a neighbourhood $\mathcal{N}$ of 3×31 pixels was used. In Fig. 6 the results are shown for the basic algorithm without the scaling parameter (top) and the extended algorithm that search for the optimal scaling (bottom). The lines at the back of the graphs show the mean value.

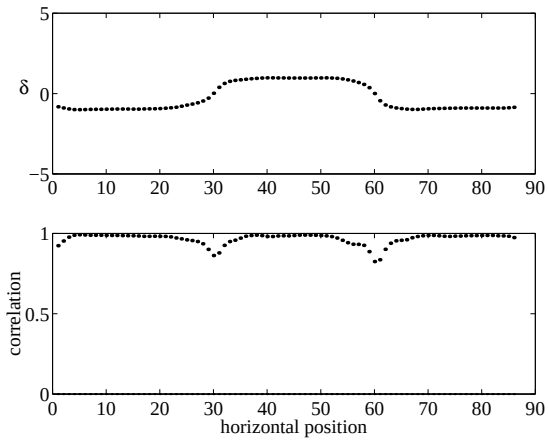

Fig. 5. Top: line 38 from the disparity estimates in Fig. 3. The small dots indicates the disparity estimates with the second strongest correlations. Bottom: The corresponding correlations.

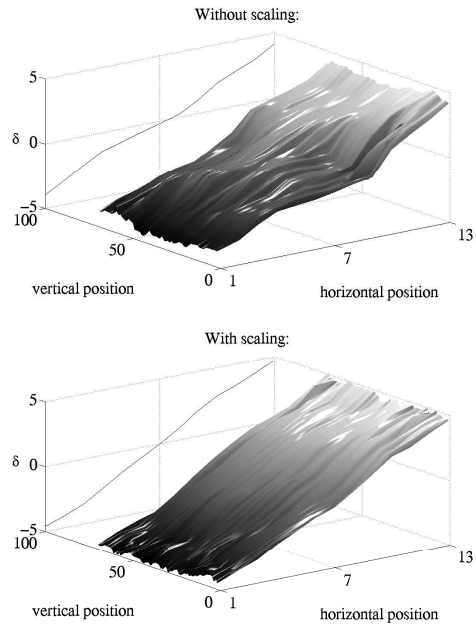

Fig. 6. Disparity estimate for a scale difference of one octave between the images without scale analysis (top) and with scale analysis (bottom).

The filters used created by the CCA is illustrated in Fig. 7 in the spatial domain and Fig. 8 in the frequency domain.

The third experiment illustrates the algorithm's capability of multiple disparity estimates on semi-transparent images. The test images in this experiment were generated as a sum of two

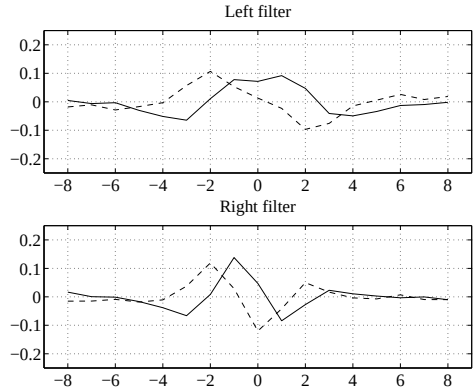

Fig. 7. The filter created by CCA in the spatial domain. Solid lines show the real parts and dashed lines show the imaginary parts.

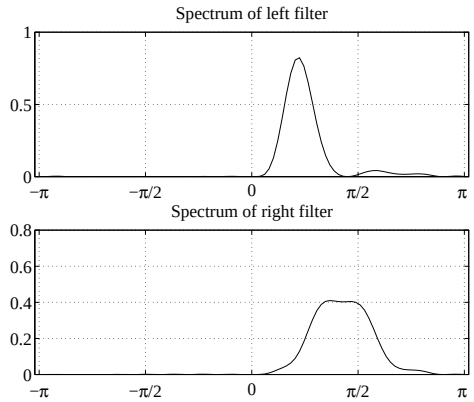

Fig. 8. The filter created by CCA in the frequency domain.

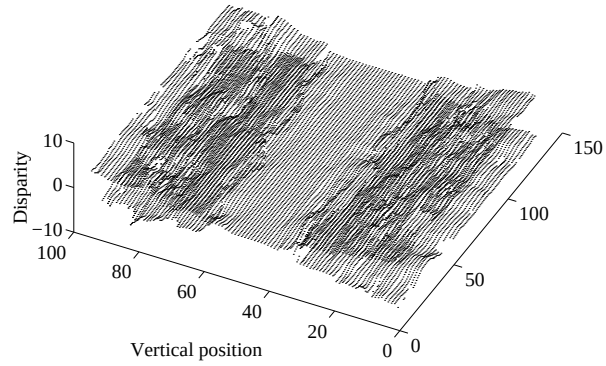

Fig. 10. The result for the semi-transparent images.

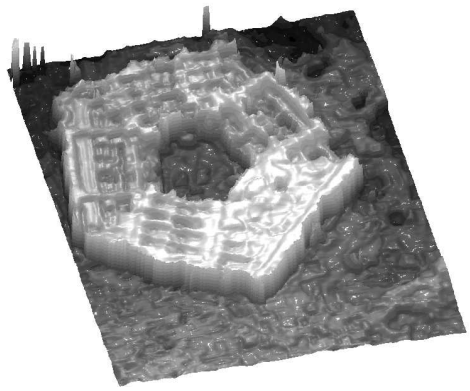

Fig. 12. Result for the pentagon images.

images with white uncorrelated noise. The images were tilted in opposite directions around the horizontal axis. The disparity range was $+/-5$ pixels. Fig. 9 illustrates the test scene. A neighbourhood $\mathcal{N}$ of 31×3 pixels was used for the CCA. The result is shown in Fig. 10. The result show that the disparities of both the planes are approximately estimated. In the middle, where the disparity difference is small, the result is an average between the two disparities in accordance with the results illustrated in figure 5.

The final experiment, illustrates how the algorithm works on a real stereo image pair. The stereo pair is two air photographs of Pentagon (see Fig. 11). A neighbourhood $\mathcal{N}$ of 7×7 pixels was used in this experiment. The result is shown in Fig. 12.

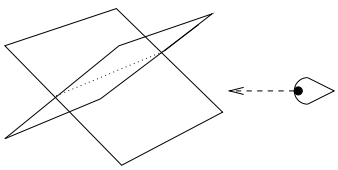

Fig. 9. The test image scene for semi-transparent images.

Note the discontinuities in the depth estimates at the walls.

VI. Conclusions and discussion

We have presented a novel stereo algorithm with sub-pixel accuracy that can handle depth discontinuities and different scalings of the images. The algorithm combines canonical correlation analysis and phase analysis. The algorithm can efficiently handle scale differences of up to one octave between the images. Scale differences occur when the image patch is not perpendicular to the observer. So far we have only used a basis filter set of two identical filters shifted two pixels. A larger filter set can be used which can contain both filters with different spatial positions and filters with other frequency functions. Such a set would allow for a wider range of disparities and scales to be analysed.

Another natural extension of the algorithm is to include also vertical shifts. Such an extended algorithm could for example be used for motion detection.

References

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, second edition, 1984.

[2] S. T. Barnard and M. A. Fichsler. Computational Stereo. *ACM Comput. Surv.*, 14:553–572, 1982.

Fig. 11. Stereo pair of Pentagon.

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

[4] S. Becker and G. E. Hinton. Learning mixture models of spatial coherence. *Neural Computation*, 5(2):267–277, March 1993.

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

[6] J. S. Bendat and A. G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. John Wiley & Sons, 1980.

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

[8] 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.

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

[10] A. D. Jepson and D. J. Fleet. Scale-space singularities. In O. Faugeras, editor, *Computer Vision-ECCV90*, pages 50–55. Springer-Verlag, 1990.

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

[12] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

[13] 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.

[14] D. Marr. *Vision*. W. H. Freeman and Company, New York, 1982.

[15] E. Oja. A simplified neuron model as a principal component analyzer. *J. Math. Biology*, 15:267–273, 1982.

[16] T. D. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 59:405–418, 1988.

[17] C. E. Shannon. A mahtematical 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.

[18] C-J. Westelius. *Focus of Attention and Gaze Control for Robot Vision*. PhD thesis, Linköping University, Sweden, S–581 83 Linköping, Sweden, 1995. Dissertation No 379, ISBN 91–7871–530–X.

[19] R. Wilson and H. Knutsson. A multiresolution stereopsis algorithm based on the Gabor representation. In *3rd International Conference on Image Processing and Its Applications*, pages 19–22, Warwick, Great Britain, July 1989. IEE. ISBN 0 85296382 3 ISSN 0537–9989.