INDEPENDENT COMPONENTS OF OPTICAL FLOWS HAVE MSTd-LIKE RECEPTIVE FIELDS

Ki-Young Park[†], *Marwan Jabri*[‡], *Soo-Young Lee*[†] *and Terrence J. Sejnowski*[§]

†Brain Science Research Center and also Department of Electrical Engineering
Korea Advanced Institute of Science and Technology, Taejeon, Korea
‡Computer Engineering Laboratory, School of Electrical and Information Engineering
The University of Sydney, Sydney, Australia
§Computational Neurobiology Laboratory, The Salk Institute, California, USA

ABSTRACT

We describe in this paper the properties of independent components of optical flow of moving objects. Video sequences of objects seen by an observer moving at various angles, directions and distances are used to produce optical flow maps. These maps are then processed using independent component analysis which yields filters that resemble the receptive fields of dorsal medial superior temporal cells of the primate brain. Contraction, expansion, rotation and translation receptive fields have been identified. Our results support Barlow's sensory coding theory and is in-line with other work that dealt with the independent components of image and video intensities.

1. INTRODUCTION

Barlow proposed about fourty years ago that the brain could represent sensory information using factorial code [1]. More recently researchers have reported the emergence of independent components from natural images [2] and video sequences [3] when entropy maximization techniques are used on their intensities. The independent components of natural images have properties similar to the localized edge receptive fields of simple cells in the primary visual cortex of mammals. Those of video sequences resemble localized spatiotemporal receptive fields – moving edge filters [4].

It is known that complex visual motion processing is performed by the middle temporal (MT) and medial superior temporal (MST) areas. In particular the dorsal region of MST (MSTd) has attracted a great deal of neurophysiological interest because of its role in processing complex visual motion patterns. Cells in this area have large receptive fields and respond selectively to the expansion, rotation, and spiral motion stimuli that are generated during observer motion.

Area MSTd receives its primary input from the MT area. MT cells produce highly selective responses to directional motion and speed in their relatively small regions of the visual field. Hence it is considered their activities represent optical flow information, though representation aspects are not well understood. MT cells also respond to motion disparity.

If we extend the factorial representation/coding hypothesis to MT and MST, the question is what would be independent components of complex motions and how well they would fit to the properties of the receptive fields of MSTd cells.

Zemel and Sejnowski [5] hypothesized complex optical flows produced by the combination of observer motion with other independently moving objects were composed of multiple regular patterns to which MSTd cells had been found to be selectively tuned and they suggested a functional role for the MST area: to encode the ensemble of motion causes that generated the complex flow field. They proposed an MST model based upon an auto-encoder neural network, which was trained by a cross-entropy measure. Following training, the filters of the hidden layer of the auto-encoder were found to selectively respond to specific motion like rotation, contraction and translation.

We describe in this paper simulation experiments aimed at exploring the properties of independent components extracted from optical flow of complex motion. We have used a ray tracing system to generate video sequences of objects seen by an observer moving at varying distances, angles and directions. The optical flow of the sequences was computed and independent components were extracted using an ICA algorithm. Our hypothesis was that filters produced using the analysis would have similarities with the receptive fields of MSTd cells. An analysis of the independent components revealed this similarity and filters tuned to contraction, rotation and translation were present. Our results support the theory of sensory coding proposed by Barlow and elaborated by others [6, 7]

2. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis (ICA) is an information maximisation method for extracting the causes or sources from multidimensional observations. ICA has been applied to blind source separation and feature extraction problems.

ICA has been applied successfully to the extraction of edges from still natural scene images [2] and spatiotemporal edges from the intensities of video sequences [3].

In studies on feature extraction, observations like natural images are assumed to be linear combinations of several underlying basis functions represented by the columns of a matrix $\mathbf{A}$ and an underlying vector of 'cause', $\mathbf{s}$ [8, 9]. Each element of vector $\mathbf{s}$ has its own associated basis funtion, and represents an underlying 'cause' of the image. The linear image synthesis model is described by:

$$\begin{aligned} \mathbf{x} &= \sum_i s_i \mathbf{a}_i \\ &= \mathbf{A}\mathbf{s} \end{aligned} \tag{1}$$

where $\mathbf{x}$ is observation vector and $\mathbf{a}_i$ is a column of matrix $\mathbf{A}$.

The underlying causes can be extracted by corresponding independent component filters $\mathbf{w}_i$ which constitutes the rows of $\mathbf{W}$.

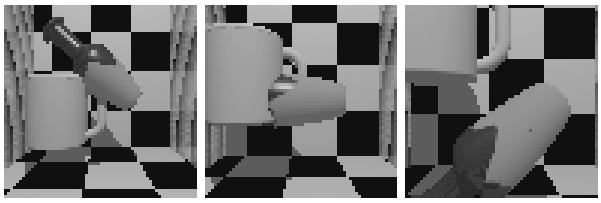
$$u_i = \mathbf{w}_i \cdot \mathbf{x} \tag{2}$$

where $\cdot$ operation denotes the inner product and u_i is an element of the recovered underlying cause $\mathbf{u}$ which responds to a specific feature of observation $\mathbf{x}$ captured by the related filter $\mathbf{w}_i$.

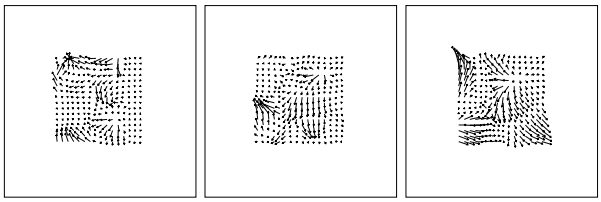
The problem is to find $\mathbf{W}$, if it exists, with the assumption that underyling causes are statistically independent and this can be done by ICA algorithms. In the simulations we report below we have used the ICA learning rule which maximizes the entropy of output $\mathbf{y}$, the nonlinear transformed version of $\mathbf{u}$ as reported by Bell and Sejnowski [2].

3. OPTICAL FLOW

Optical flow is an approximation to the 2-d motion field – a projection of the 3-d velocities of surface points onto the imaging surface – from spatiotemporal patterns of image intensity [10]. But computing two components of optical flow at each pixel is an ill-posed problem due to the intrinsic lack of constraints. Most algorithms that use temporal derivatives to estimate the varying image intensity fail to extract exact optical flow maps for the images which contain occlusions of objects, since they assume smooth changes of intensities.



(a) video frame



(b) optical flows

Figure 1: Extract of an example video sequence and its associated optical flow

In order to deal with object occlusions, Nagel used second-order derivatives to measure optical flow and suggested oriented-smoothness constraints where smoothness is not imposed across steep intensity gradients (edges). This prevents smoothing over intensity discontinuities [11, 10] most likely to represent object boundaries. Since our video sequences have many occlusions, and because occlusions represent important information for motion segmentation, we selected Nagel's algorithm for optical flow computation.

In contrast with image intensities, each pixel in an optical flows map has two components (called x and y here) and thus a coding scheme is needed to apply independent component analysis. The representation we have used is the simplest. For each optical flow observation, we concatenated of all x components followed by all y components.

4. EXPERIMENTS

4.1. Methods

All video sequences used in our simulations were synthetic scenes created by a computer program (Persistence of Vision Ray Tracer, which is publically available at http://www.povray.org). This program allows simu-

lation of dynamic scenes including various stationary or moving objects, backgrounds and observer motions. A scene contains upto three objects from a choice of seven types of objects – ball, cup, table, chair, cube, vase and table lamp – and variouse backgrounds (texture) which can be composed of planes of 5 different patterns. Figure 1 shows an extract of an example video sequence and the computed optical flow.

For each video sequence, a virtual space with a background was defined and the number of objects was set, together with an initial and final positions of an observer. Each object was placed at random in the space. The sequence was generated as follows:

1. The field of view was 60×60 deg.

2. The observer was stationary with a probability of $1/3$; The observer moved in x-direction or z direction with probabilities of $1/3$ and $2/3$ respectively; The observer could rotate to track a point or an object.

3. An object could move in all directions independently with a probability $1/5$ in each of the x, y and z directions. Hence about half of the objects were in motion and any (complete or partial) occlusion of objects was allowed. In addition to translation, an object could rotate independently in the x-y plane.

For each sequence the generation parameters were determined and a script was produced to generate a video sequence. A total of 30 frames was produced by specifying 3-D positions of the camera and each object in the image and then updating the pose of the camera and each object based on the motion parameters. The script contained the description of the content of an image and called a set of graphics routines supplied in the ray-tracing package to render each image. The script also used a library contained in the package that included descriptions of the seven object shapes and backgrounds.

A total of 15,000 movies were created. From these, 45,000 optical flow maps were extracted for training. To produce an optical flow map, 15 frames were used through gaussian smoothing. The size of each frame was 64×64 which corresponds to a 60×60 deg visual field and due to the spatial smoothing and a 2-to-1 subsampling, the size of the resulting optical flow map was 16×16. Hence, the training data consisted of 45,000 vectors of length 512 (optical flow map is 16×16, and each element has 2 dimensions). These vectors were zero-meaned and sphered before being used by the ICA algorithm.

We have used the Matlab ICA toolbox written by Scott Makeig and colleagues, which implements the information maximization algorithm of Bell and Sejnowski with the natural gradient feature of Amari, Cichocki and Yang [12, 2, 13].

The hyperbolic tangent funtion was used as the nonlinearity and no bias term was used. We have used PCA preprocessing for dimension reduction prior to applying ICA. PCA is typically used to reduce dimension of input data based on second order statistics.

4.2. Results

Figure 2 shows some of the filters extracted by ICA. ICA was performed both with and without PCA as a preprocessing step.

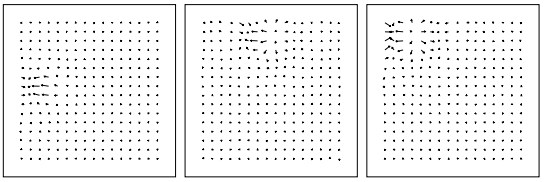
In cases where PCA was performed before ICA, the yielded filters were meaningful. However without PCA, the filters had too small receptive regions and showed no regular patterns. This means that the number of underlying basis functions to produce optical flow of dynamic scenes was quite small compared to the dimensions of input data (512 here). Using PCA, we reduced the dimensions of the input data to 50, 70, 100 and 200. More than 200 principal components seemed to be too many and less than 50 to be too small. The number of pricipal components between 50 and 150 qualitatively resulted in similar filters.

The ICA filters shown in figure 2 correspond to a PCA preprocessing reducing the dimensions down to 100. The filters in the top row are receptive fields that respond to contration and expansion, the middle row correspond to rotation and the third row corresponds to translation.

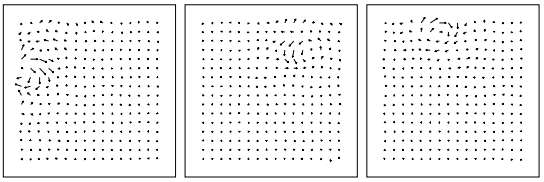
Although the quantitative measure was not made as yet for the selectivity of these units, each filter can be predicted to be spatially localized and selectively responds to a specific type of motion.

The filters shown in the bottom row of figure 2 can be considered to respond to the combination of more than two types of movements. This could be similar to the response of MSTd cells that have been observed in many studies and that Duffy and Wurtz call double-, triple-component neurons [14].

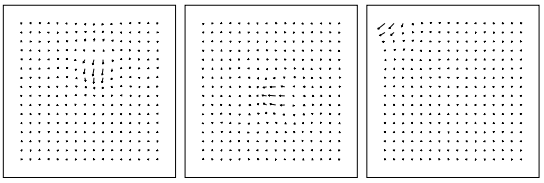
To test the reponse patterns of each filter, 15,000 optical flows were generated from 5,000 video sequences which were produced using the same procedures as the training set. The response pattern of each filter for these test data are shown in figure 3. Figure 3(a) shows the number of output units (filters) which give the output activation of each range. Outputs of all units were linerly normalized to have the same maximum value over all input data and trivial input stimulus which



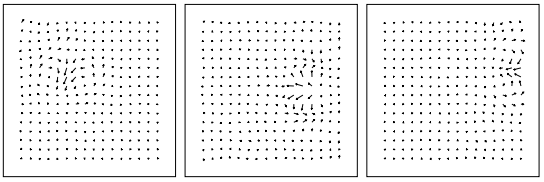
(a) Expansion/Contraction



(b) Rotation

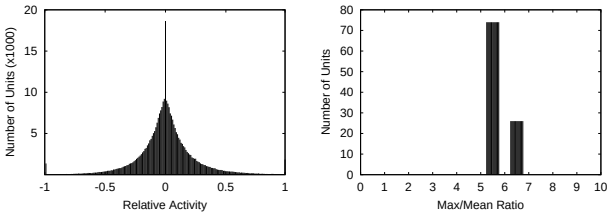


(c) Translation



(d) Combination

Figure 2: Examples of filters extracted by ICA



(a) Sparseness (b) Selectivity

Figure 3: Distribution of output amplitude for test patterns. The distribution of amplitude shows the sparseness of output activation.

contains little movememt were discarded. This normalization was to prevent redundant input which made all output units inactive. As shown in the figure, due to the learning constraints of ICA and the assumed distribution of underlying sources, the response patterns shows sparseness of filters – only a small number of units respond for a given input pattern.

Figure 3(b) shows the selectivity of the filters. Since a filter responds only to preferred types of input and is inactive for all others patterns, the ratio between the maximum activity of an output unit and its mean over the testing set is quite large.

5. DISCUSSION AND FUTURE WORK

In this study, independent component analysis was performed on the optical flow of complex motion and the resulting filters show characteristics that ressemble those of MSTd cells.

The resulting filters were tuned to specific motion patterns, were moderate in size and were localized in their positions. Filters selective to translations, rotations, contractions and expansions have been observed.

However some important properties of MSTd cells could not be observed explicitely from our present results. Firstly, some studies on MSTd reported cells with some levels of position invariance. The linear filter model described in this paper could not show any invariance. By introducing interactions between these linear filters, it is possible that position invariance can be achieved to some extent [15, 5].

Secondly, more laborate modeling of MT cells is required. Many units were selective to vertical or horizontal tranlational movements and little were selective to translation at arbitrary angles. This is probably due to the fact that in the coordinate system we adopted any movement could be represented as linear combination of vertical and horizontal movements. A more

elaborate model of MT cells like that of Nowlan and Sejnowski [16] and that represents motion components by population coding of many direction selective units may be able to produce translation selective units at various angles. An alternative representation is to use a log-Polar coordinate system similar to that used by Grossberg and colleagues [17].

6. REFERENCES

[1] H. B. Barlow. Possible principles underlying the transformation of sensory messages. In W.A. Rosenbluth, editor, *Sensory Communication*, pages 371–394. MIT Press, 1961.

[2] A. J. Bell and T. J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[3] J. H. van Hateren and A. van der schaaf. Independent component filters of natural images compared wiht simple cells in primary visual cortex. *Proceedings of Royal Society of London B*, 265:2315–2320, 1998.

[4] J. H. van Hateren and D. L. Ruderman. Independent component analysis of natural image sequences yield spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of Royal Society of London B*, 265:2315–2320, 1998.

[5] R. S. Zemel and T. J. Sejnowski. A model for encoding multiple object motions and self-motion in area MST of primate visual cortex. *The Journal of Neuroscience*, 18(1):531–547, 1998.

[6] D. J. Field. What is the goal of the sensory coding? *Neural Computation*, 6:559–601, 1994.

[7] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society America A*, 4:2379–2394, 1987.

[8] A. J. Bell and T. J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.

[9] B. A. Olshausen and D. J. Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7(2):333–339, 1996.

[10] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.

[11] H.-H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.

[12] S. Makeig. ICA toolbox for psychophysiological research (version 3.4). WWW Site, Computational Neurobiology Laboratory, The Salk Institute for Biological Studies `<www.cnl.salk.edu/~ica.html>` [World Wide Web Publication], 1999.

[13] S. Amari, A. Cichocki, and H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing System 8*, pages 757–763, 1996.

[14] C. J. Duffy and R. H. Wurtz. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selecitivity to large-field stimuli. *Journal of Neurophysiology*, 65(6):1329–1345, 1991.

[15] K. Zhang, M. I. Sereno, and M. E. Sereno. Emergence of position-independent detectors of sense of rotation and dilation with Hebbian learning: An analysis. *Neural Computation*, 5(4):597–612, 1993.

[16] S. J. Nowlan and T. J. Sejnowski. A selection model for motion processing in area MT of primates. *The Journal of Neuroscience*, 15(2):1195–1214, 1995.

[17] S. Grossberg, E. Mingolla, and C. Pack. A neural model of motion processing and visual navigation by cortical area MST. *Cerebral Cortex*, 9:878–895, 1999.