

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 optic flow of moving objects. Video sequences of objects seen by an observer moving at various angles, directions and distances are used to produce optic flow maps. These maps are then processed using independent components analysis which yields filters that resembles 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 maximisation techniques are used on their intensities. The independent components of natural images have properties similar to the localised edge receptive fields of simple cells in the primary visual cortex of mammals. Those of video sequences resemble localised 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 (dMST) 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 dMST 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 optic 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 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 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 [5].

We describe in this paper simulation experiments aimed at exploring the properties of independent components extracted from optic 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 optic flow of the sequences was computed and independent components were extracted using ICA algorithms. Our hypothesis was that filters produced using the analysis would have similarities with the receptive fields of dMST cells. An analysis of the independent components reveals this similarity and filters tuned to contraction, rotation and translation were present. Our results sup-

port the theory of sensory coding proposed by Barlow and elaborated by others [6, 7]

2. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis(ICA) is a statistical 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 and spatiotemporal edges from still natural scene images [2] and 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 function, and represents an underlying ‘cause’ of the image. The linear image synthesis model is therefore:

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

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

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

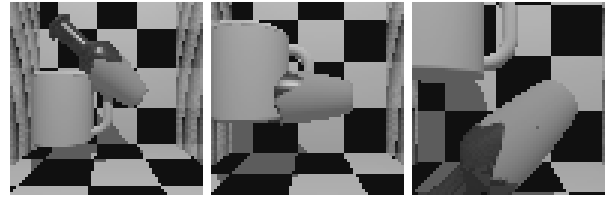
$$u_i = \mathbf{w}_i \cdot \mathbf{x} \quad (2)$$

where \cdot operation means 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 .

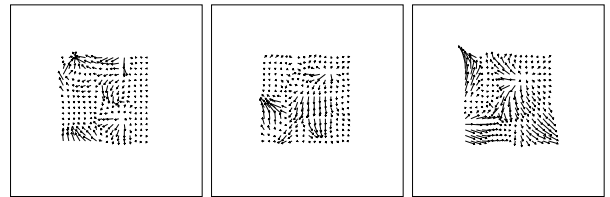
The problem is to find \mathbf{W} if it exists and with the assumption that underlying causes are statistically independent and this can be done by the ICA algorithm. In the simulations we report below we have used the ICA learning rule which maximizes the entropy of output \mathbf{y} , the nonlinear 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



(a) video frame



(b) optical flows

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

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 derivative to estimate the varying image intensity fail to extract exact optical flow maps for the images which contain the occlusions of objects, since they assume smooth changes of intensities.

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 the segmentation of motion, we selected Nagel’s algorithm for optic flow computation.

In contrast with image intensities, each pixel in an optical flows map has two components (we call X and y) and thus a coding scheme is needed to apply independent component analysis. The representation we have used is the simplest. For each optic 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 simulation of dynamic scenes including various stationary or moving objects, backgrounds and observer motions. A scene contains upto three objects from a choice of 7 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 optic 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 position of an observer. Each object was place 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 , z direction. Hence about half of the objects were in motion and any occlusion or overlapping 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 scripirt was produced to generate the video sequence. A total of 20 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 six objects and the backgrounds.

A total of 15,000 movies were created. From these, 45,000 optic flow maps were extracted for training. To produce an optic 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 optic flow map was 16×16 . Hence, the training data consisted of 45,000 vectors of length 512 (optic flow map is 16×16 , and each componenet has 2 dimensions). These vectors were zero-meaned and sphered before being used by the ICA algrotihms.

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, Cichoci and Yang [12, 2, 13].

The hyperbolic tangent funtion was used as the non-linearity 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.

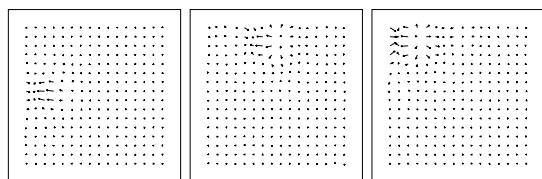
In cases where PCA was performed before ICA, the yielded filters were meaningful. However without PCA, the filter had too small receptive regions and showed no regular patterns. This means the number of the underlying basis function to produce optic flow of dynamic scenes was quite small compared to the dimension of input data (512 here). Using PCA, we reduced the dimension 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 PCs 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.

Three filters in the top row are receptive fields that respond to contraction and expansion, the middle row correspond to rotation and the third row corresponds to translation.

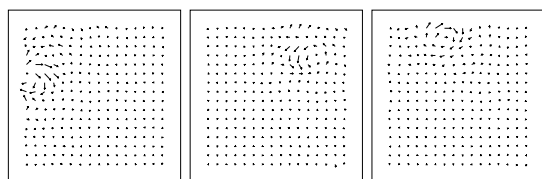
Although the quantative measure was not made as yet for the selectivity of these units, each filter can be predicted to be spatially localized and selectively responds to a specific type of motion.

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

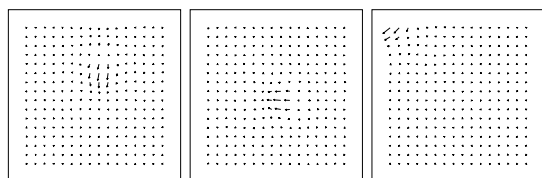
To test the reponse patterns of each filter, 15,000 optic flows were generated from 5,000 video sequences



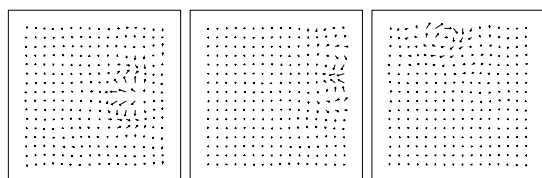
(a) Expansion/Contraction



(b) Rotation

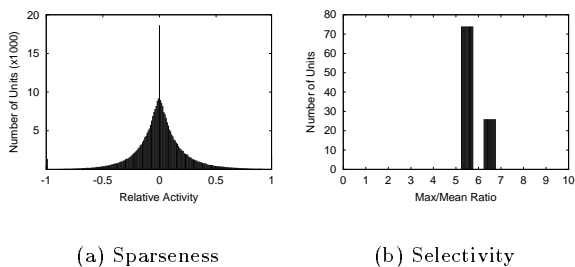


(c) Translation



(d) Combination

Figure 2: Examples of filters constructed 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.

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 linearly normalized to have the same maximum value over all input data and trivial input stimulus which contains little movement 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 optic flow of complex motion and the resulting filters show characteristics that resemble those of MSTd cells.

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

However some important properties of MSTd cells are missing or could not be observed explicitly from our results. Firstly, some studies on MSTd reported cells with some levels of position invariance. The linear property of the model in this paper could not show any invariance.

Many units were selective to vertical or horizontal translational 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.

Secondly, more elaborate modeling of MT cells which represent motion components by population coding of many direction selective unit like that of Nowlan and Sejnowski will be able to produce filters of translation selective units at various angles [15]. An alternative is to use a log-Polar coordinate system similar to that reported by Grossberg and colleagues [16].

6. REFERENCES

- [1] H. B. Barlow. Possible principles underlying the transformation of sensory messages. In W.A. Rosenbluth, editor, *Sensory Communication*, number 371–394. MIT Press, 1961.
- [2] Anthony J. Bell and Terrence 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 with simple cells in primary visual cortex. *Proceedings of Royal Society 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 London B*, 265:2315–2320, 1998.
- [5] Richard S. Zemel and Terrence 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] David J. Field. What is the goal of the sensory coding? *Neural Computation*, 6:559–601, 1994.
- [7] David 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] Anthony J. Bell and Terrence J. Sejnowski. The ‘independent components’ of natural scenes are edge filters. *Vision Research*, 37(23), 1997.
- [9] B. A. Olshausen and D. J. Field. Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, 7(2), 1996.
- [10] D. J. Fleet, J. L. Barron and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.
- [11] Hans-Hellmut Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, 33:299–324, 1987.
- [12] Scott et al. Makeig. ICA toolbox for psychophysiological research (version 3.2), 1998.
- [13] A. Cichoci S. Amari and H. Yang. A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing System 8*, pages 757–763, 1996.
- [14] Charles J. Duffy and Robert H. Wurtz. Sensitivity of MST neurons to optic flow stimuli. i. a continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, 65(6):1329–1345, 1991.
- [15] Steven J. Nowlan and Terrence J. Sejnowski. A selection model for motion processing in area mt of primates. *The Journal of Neuroscience*, 15(2):1195–1214, 1995.
- [16] E. Mingolla S. Grossberg and C. Pack. A neural model of motion processing and visual navigation by cortical area mst. *Cerebral Cortex*, 9:878–895, 1999.