Computational Model for Rotation-Invariant Perception

Wenlu Yang[1,2], Liqing Zhang[1]*and Libo Ma[1]
Department of Computer Science and Engineering, 800 Dongchuan Rd.,
Shanghai Jiaotong University, Shanghai 200240, China
Department of Electronic Engineering, 1550 Pudong Rd.
Shanghai Maritime University, Shanghai 200135, China
wenluyang@online.sh.cn; lqzhang@sjtu.edu.cn

Abstract

Visual perception of rotation is one of important functions of processing information in the visual pathway. To simulate the mechanism, we propose a model for perception of rotation. First, we briefly introduce the rotation-invariant basis functions learned from natural scenes using Independent Component Analysis (ICA). We used these basis functions to construct the perceptual model. By using the correlation coefficients of two neural responses as the measure of rotation-invariance, our model can perform the task of perception of rotating angles. Computer simulation results show that the present model is able to perceive rotation-invariance and successfully perceive the relative angles of rotating patches.

1. Introduction

We can recognize an object regardless of its distance, position or rotating angle. In the mathematical term, object recognition is not influenced by its transformation, such as translation, rotation or scale. Such capability of our recognizing transformation-invariant objects can become an inherent ability whether it is native or of learning in our infancy. Many recent researches in the fields of neuroscience, neurophysiology and psychology show that such a transformation invariant preprocessing could be a necessary step to achieve transformation-invariant classification or detection in a hierarchical system. In this paper, we will focus on the mechanism of rotation invariance. We will propose a hierarchical model that simulates the mechanism in the visual pathway. On the other hand, due to evolution from nature in the long term, this mechanism has an important correlation with statistical properties of natural scenes. Following the way, Barlow[1, 2] found that the role of early sensory neurons in the visual pathway is to remove statistical redundancy in the sensory inputs, suggesting that Redundancy Reduction is an important processing principle in the neural system. Based on this concept, Gabor-like features resembling the receptive fields of V1 cells have been derived either by imposing sparse over-complete representations[3] or statistical independence as in independent component analysis[4].

However, these studies have not taken transformation invariance into account, and the question is how well this line of research predicts the full spatiotemporal receptive fields of simple cells. For example, when an image rotates within receptive fields of simple cells, how do the simple cells and complex cells response? Some researchers have investigated the question. Hyvarinen and Hoyer[5, 6] modelled receptive fields of complex cells and Van Hateren [7]obtained spatiotemporal receptive fields of complex cells. Grimes and Rao[10] proposed a bilinear generative model to study the translation-invariance. However, there are few models in the literatures of physiologically and neurophysiologically perceiving rotation of objects or images. To investigate the question, we apply Independent Component Analysis(ICA) to learning from natural scenes the rotation-invariant features, and use the features to construct a model for rotation-invariant perception. The goal of the model is to perceive the rotation of patches from natural images.

The rest of the paper is organized as follows. Section 2 introduces our model for perception of rotation invariance. In section 3, we will demonstrate perception of rotation-invariance and rotation of patches. In the final section we provide some discussions and conclusions.

2. Model for perception of rotation invariance

In this section, we first show rotation-invariant basis functions which can be learned from natural scenes. Then using the basis functions, we propose a computational model for perception of rotation invariance.

*To whom correspondence should be addressed.

2.1. Rotation-invariant basis functions

To learn rotation-invariant basis functions from natural scenes, we sample small patches from a set of big natural images. A sampling window with size of 11×11 is randomly located on a big natural scene and a patch with the same size of the sampling window is selected, denotes u_1. Then fix the same center, clockwise rotate the sampling window by an interval of 15 degree, another patch is sampled. Again, rotate the window and sample the next one, till twenty-four times. In the same way, the total twenty-four patches are sampled and then reshaped to one column vector as a sample, size of 2904-by-1. From the set of big natural images, we obtain total 20000 of samples as the training data for learning rotation-invariant basis functions. Applying ICA with the learning algorithm based on the Natural Gradient[8, 9] to the training data yields the rotation-invariant basis functions, shown in Figure 1. Any of basis functions resembles the receptive field of a simple cell in V1, and a group of basis functions in a row is similar to the receptive field of a complex cell which performs perception of rotation invariance. The neighboring basis functions in a row have an interval of fifteen degree of counter-clockwise rotation. Basis functions in a column are in the same group of which elements are used to reconstruct the input patterns while given corresponding activities of simple cells. For the convenience of viewing the regularity, these basis functions are arranged in counter-clockwise along the circumference with an interval of fifteen degree, shown in Figure 2. Every circle has the properties of the receptive field of a complex cell.

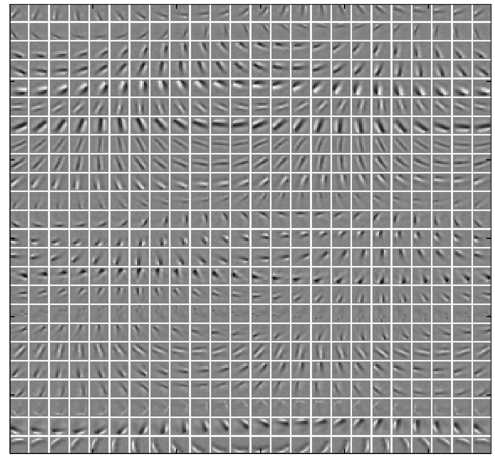

Figure 1. Subsets of rotation-invariant basis functions. Basis functions in a row resemble the receptive field of a complex cell which performs perception of rotation invariance. Rotational interval is fifteen degree.

2.2. The perception model

In this section, we will present a model for rotation-invariant perception, shown in Figure 3. The input patterns are two patches with parameters of α_i and α_j, respectively. Here, α_i and α_j are the rotating angles within the range of zero and three hundred sixty degree by an interval of fifteen.

The middle layer of the model is a group of rotational basis functions through which neurons respond to input patterns. These basis functions are arranged in a matrix. There are 100 basis functions in one row and 24 in one column. Those in one row have different parameters of characteristics: location, orientation and bandpass. These characteristics inhere in the receptive fields of simple cells. And those in one column have some variant parameters which are obtained by rotating their neighboring basis functions.

In the next layer of the model, there are two groups of neurons. One group of neurons receives the stimulus $u^{t_1}_{\alpha_i}$ at time t_1 and the other group receives the stimulus $u^{t_1}_{\alpha_j}$ at time t_2. Where, one of neurons in a row fires while the stimulus contains the content of the same orientation as that of receptive field of the neuron. If the content appears at time t_1 and its rotation transformation at time t_2, there are two neurons which fire at time t_1 and t_2, respectively. And that, the two neurons must be at the same column because that the corresponding basis functions of the two neurons can be transformed to each other by rotation transformaton.

After neurons receiving the stimuli u_{α_i} at time t_1 and u_{α_j} at time t_2 , the final layer of the model calculates the correlation coefficients between any two responses $\mathbf{X}^{t_1}_{\alpha_i}(i = 1, ..., M)$ and $\mathbf{X}^{t_2}_{\alpha_j}(j = 1, ..., M)$. The max coefficient is selected to determine the relative value of rotating angles. The index (i, j) of the maximum in coefficient matrix will tell us the relative difference such as counter-clockwise rotation angle $\Delta\theta$. It is necessary to note that we only need the relative transformation, not the absolute value of parameters of the stimuli. If $j \geq i$, $\Delta\theta = (j - i) \times 360/M$; otherwise, $\Delta\theta = (M + j - i) \times 360/M$. An example of rotation perception is demonstrated in section 3.2.

3. Simulations and Results

We will show two experimental results to verify the performance of our proposed model. One is how the neurons with different rotation angles respond when twenty-four patches are feed to the model. These twenty-four patches are transformed to each other by rotation transformation. The another demonstrates how to perceive rotation of image patches. And we will present perception of rotating faces as

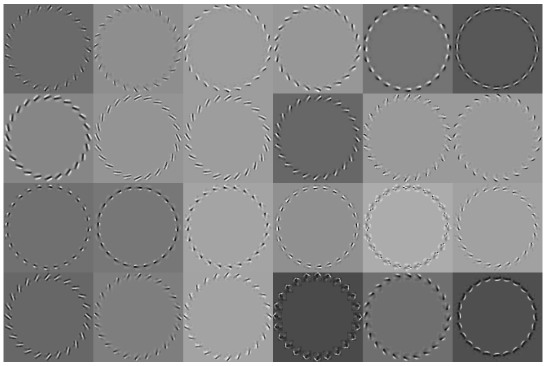

Figure 2. Subsets of rotation-invariant basis functions. Every circle contains twenty-four basis functions which are arranged in counter-clockwise along the circumference with an interval of fifteen degree. Every circle resembles the receptive field of a complex cell which is able to perceive rotation invariance. The first circle is the rearrangement of the basis functions in the first row in Figure 1.

an application of the model.

3.1. Rotation invariance

In this section, we will investigate the rotation invariance. The rotation invariance means that the response of a neuron slightly fluctuates while the stimulus rotates within the receptive field of the neuron. First we generate a testing sample U which is composed of twenty-four patches $\{u_{\alpha_i}\}(i = 1, \cdots, 24)$ selected from natural images by the method mentioned in section 2.2. The Sequences $\{u_{\alpha_1}, u_{\alpha_2}, \cdots, u_{\alpha_{24}}\}$, as input patterns shown in Figure 4, are feed to the perceptual model. The corresponding responses of neurons through the basis functions $\{F_{\alpha_1}, F_{\alpha_2}, \cdots, F_{\alpha_{24}}\}$ are plotted in Figure 5. From the Figure 5, the similarity of responses of rotation-invariant neurons is very high while the stimulus rotates within their receptive fields.

Randomly select two responses, for example, $X^{t_1}_{\alpha_6}$ and $X^{t_1}_{\alpha_{11}}$, examine the dispersion between them, as shown in Figure 6. The bottom line tells us that the dispersion between $X^{t_1}_{\alpha_6}$ and $X^{t_1}_{\alpha_{11}}$ is very small. In other words, the rotation-invariant neurons activate while the stimulus rotates within their receptive fields. These neurons are able to perceive rotation invariance.

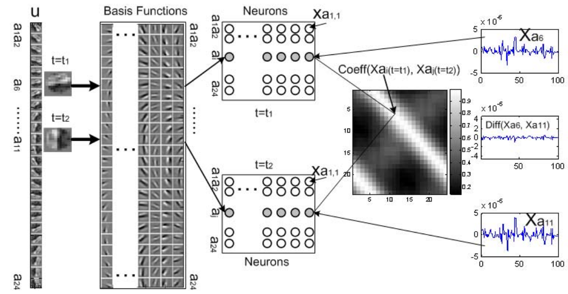

Figure 3. Model for transformation-invariant perception. The input patterns are the rotation-transformed data. $x^{t_1}_{\alpha_{i,k}}(k = 1, 2, \cdots, N)$ denotes the response of the k-th neuron in the row α_i responding to stimulus u_{α_i} at time t_1 through the basis function $F_{\alpha_{i,k}}$. And so does the response $x^{t_2}_{\alpha_{j,l}}(l = 1, 2, \cdots, N)$ at time t_2. $\mathbf{X}^{t_1}_{\alpha_i}$ denotes the vector of responses that the neurons in the row α_i respond to stimulus u_{α_i} at time t_1 through the subsets F_{α_i} of basis functions($F_{\alpha_i} = [F^{t_1}_{\alpha_{i,1}}, F^{t_1}_{\alpha_{i,2}}, \cdots, F^{t_1}_{\alpha_{i,N}}]^T$). That is, $\mathbf{X}^{t_1}_{\alpha_i} = [x^{t_1}_{\alpha_{i,1}}, x^{t_1}_{\alpha_{i,2}}, \cdots, x^{t_1}_{\alpha_{i,N}}]^T$.

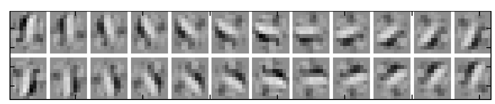

Figure 4. An example U=$\{u_{\alpha_1}, u_{\alpha_2}, \cdots, u_{\alpha_{24}}\}$ of testing patches. The corresponding rotation angles are 0, 15, $\cdots$, 345 degree from the topleft to bottomright.

3.2. Perception of rotation

After having investigated the rotation invariance, we will apply the model to perception of rotation. From the sample U, randomly select two image patches u_{α_i} and u_{α_j}(i.e. $i = 6, j = 11$). That is, the sixth and eleventh of stimuli represent rotational angles of ninety and one hundred and sixty-five in degree. The sixth patch is the first input to the perception model at time $t_1$and the eleventh at time t_2. The activities of the 24×100 neurons are plotted in Figure 7. Let $X^{t_1}_{\alpha_i}(i = 1, 2, \cdots, 24)$ denote the responses of a hundred neurons in the i-th row receiving the stimulus u_{α_i} at time t_1. Investigating carefully, we will find that $X^{t_1}_{\alpha_6}$ is similar very much to $X^{t_2}_{\alpha_{11}}$. In fact, that is the truth. Because the F_{α_6} has the same properties of transformation from the $F_{\alpha_{11}}$, and so do the stimuli u_{α_6} and $u_{\alpha_{11}}$. For more exam-

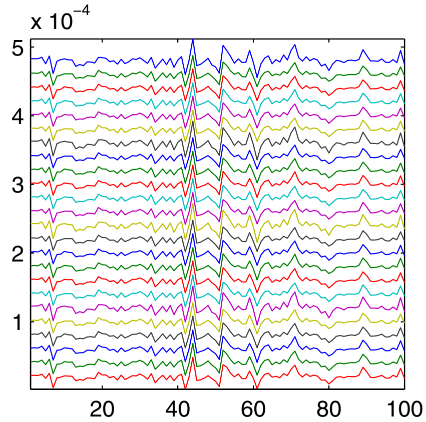

Figure 5. Responses $X_{\alpha_i}^{t_1}(i, j = 1, 2, \cdots, 24)$ of neurons with the basis functions F_{α_i}. The corresponding stimulus u_{α_i} as shown in Figure 4. The top line is the first one, and the bottom is the twenty-fourth.

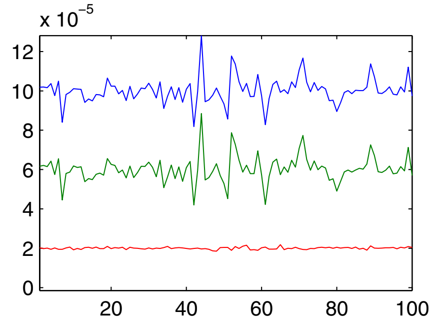

Figure 6. Rotation-invariant analysis. The top line denotes the Responses $X_{\alpha_6}^{t_1}$, the middle line denotes $X_{\alpha_{11}}^{t_1}$, and the bottom line denotes the dispersion between $X_{\alpha_6}^{t_1}$ and $X_{\alpha_{11}}^{t_1}$.

ples, $X_{\alpha_1}^{t_1}$ is similar to $X_{\alpha_6}^{t_2}$, $X_{\alpha_2}^{t_1}$ to $X_{\alpha_7}^{t_2}$, $X_{\alpha_4}^{t_1}$ to $X_{\alpha_9}^{t_2}$, and so on.

For ease of viewing the similarity of any two lines, we examine the correlation coefficients of $X_{\alpha_i}^{t_1}$ and $X_{\alpha_j}^{t_2}$. Following the idea, we compute the matrix of the correlation coefficients Coeff($X_{\alpha_i}^{t_1}, X_{\alpha_j}^{t_2}$)$(i, j = 1, 2, \cdots, 24)$, as shown in Figure 8. Find the max value from any row in the matrix and obtain its corresponding index of row and column. For instance, at the first row, the index of the max Coeff($X_{\alpha_1}^{t_1}, X_{\alpha_6}^{t_2}$) is (1,6). So, we know the relative rotating angle is (6-1)×15=75 degree according to the measure in section 2. The experimental result is the expectation of the relative angle of rotating from the stimulus u_{α_6} at time t_1 to $u_{\alpha_{11}}$ at time t_2. It is necessary to note that we only need the relative transformation, not the absolute value of angles of the stimuli.

3.3. Application for perception of face rotation

As an example of applications of the model, perception of rotating faces is given. First, use rotating faces to learn rotation-invariants face features, and then replace the basis functions in the model with the face features. When two rotating faces are feed to the model, it returns relative rotating angle. The perceptual accuracy is greater than 96%. This result is very important for face detection and recognition.

4. Discussions and Conclusions

Visual perception of rotation is one of the important functions of processing information in the visual pathway. To simulate the mechanism, we have proposed a model for perception of rotation and demonstrated some simulation results. First, we briefly introduce how to learn the rotation-invariant basis functions, and then we apply them to the perceptual model which was introduced in detail in section 2. Second, we investigated the rotation invariance of neurons in the model. Finally, we demonstrated an example of perception of rotation. Simulation results show that our proposed model successfully performs the rotation perception of rotating images. Following a rational line of application, it is of facility to apply the model to perception of the rotating velocity while given the rotating time. And it is also applied to perception of the rotating direction so long as two relative rotating angles are perceived.

However, Compared with the bilinear generative model[10] proposed by Grimes and Rao, our model has some advantages such as easy realization and less computing cost, and more transformation-invariant features. They only explored the model for learning translation-invariant basis functions, but they did not give good applications. On the basis of the idea presented in this paper, translation- or scaling-invariant basis function can be obtained and our model is easily extended for perception of translation and scaling by only simply replacing the rotation-invariant basis functions with translation- or scaling-invariant ones.

Our current efforts are focused on the perception of the rotating velocity and direction of objects in the sequences of natural videos. We are also extending the model to a framework for learning other transformation-invariant basis functions and perception of other transformations such as

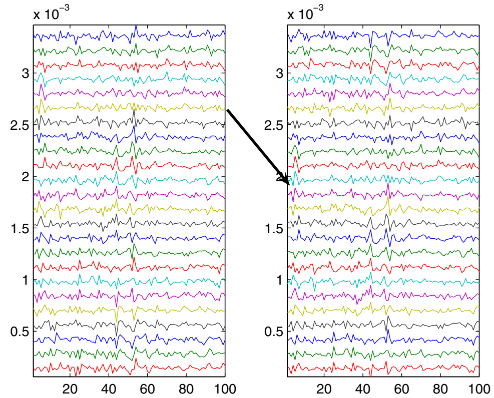

Figure 7. Responses of the neurons. The k-th line denotes the responses $X^{t_1}_{\alpha_k}$ of neurons in the k-th row in Figure 3. The top line is the first one. The left plot shows the responses to the stimulus u_{α_6} at time t_1 and the right the stimulus $u_{\alpha_{11}}$ at time t_2. For an explicit of demonstration, the responses are shifted along the Y-axis.

translation, scaling, and view changes.

5. Acknowledgment

The work was supported by the National Basic Research Program of China (Grant No. 2005CB724301) and the National High-Tech Research Program of China (Grant No.252006AA01Z125).

References

[1] Barlow H.B. Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith, editor, *Sensory Communication*, 217-234. MIT Press, 1961.

[2] Barlow H.B. Redundancy reduction revisited. *Network: Computation in Neural Systems*, Vol.12:241-253, 2001.

[3] Olshausen B.A. and Field D.J. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311-3325, 1997.

[4] Bell A.J. and Sejnowski T.J. The independent component of natural scenes are dege filters. *Vision Research*, 37:3327-3338, 1997.

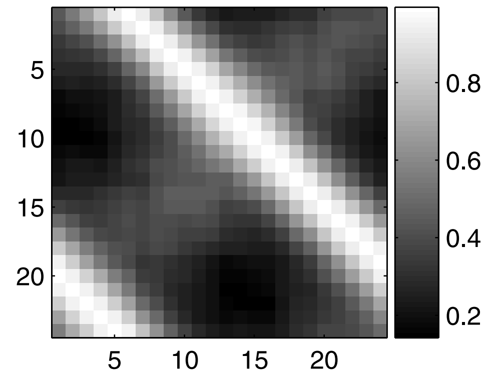

Figure 8. Correlation coefficients of responses of neurons $X^{t_1}_{\alpha_i}$ and $X^{t_2}_{\alpha_j}(i,j = 1,2,\cdots,24)$.

[5] Hyvarinen A. and Hoyer P.O. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, Vol.12(7):1705-1720, 2000.

[6] HoyerPO and Hyvarinen A. A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*, Vol.42(12): 1593-1605, 2002

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

[8] Zhang L., Cichocki A. and Amari S. Natural Gradient Algorithm to Blind Separation of Over-determined Mixture with Additive Noises, *IEEE Signal Processing Letters*, 6(11):293-295, 1999.

[9] Zhang L., Cichocki A. and Amari S. Self-Adaptive Blind Source Separation Based on Activation function Adaptation, *IEEE Transactions on Neural Networks*, 15(2):233-244, 2004

[10] Grimes, D.B. and Rao, R.P.N. Bilinear sparse coding for invariant vision. *Neural computation*, Vol.17(11):47-73, MIT Press, 2005