Perception of Transformation-Invariance in the Visual Pathway

Wenlu Yang[1,2], Liqing Zhang[1,⋆], and Libo Ma[1]

[1] Department of Computer Science and Engineering,
Shanghai Jiao Tong University, Shanghai 200240, China
wenluyang@online.sh.cn, lqzhang@sjtu.edu.cn
[2] Department of Electronic Engineering,
Shanghai Maritime University, Shanghai 200135, China

Abstract. Visual perception of transformation invariance, such as translation, rotation and scaling, is one of the important functions of processing visual information in the Brain. To simulate this perception property, we propose a computational model for perception of transformation. First, we briefly introduce the transformation-invariant basis functions learned from natural scenes using Independent Component Analysis (ICA). Then we use these basis functions to construct the perceptual model. By using the correlation coefficients of two neural responses as the measure of transformation-invariance, the model is able to perform the task of perception of transformation. Comparisons with Bilinear Sparse Coding presented by Grimes and Rao and Topo-ICA by Hayvarinen show that the proposed perceptual model has some advantages such as simple to implement and more robust to transformation invariance. Computer simulation results demonstrate that the model successfully simulates the mechanism for visual perception of transformation invariance.

1 Introduction

We can recognize an object regardless of its distance, position or rotation. In the mathematical term, object recognition is not influenced by its transformation, such as translation, rotation or scaling. 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 computational system. In this paper, we will focus on the computational mechanism for transformation invariance. We will propose a hierarchical model that simulates the mechanism in the visual pathway. On the other hand, due to biological evolution from nature in the long term, this mechanism has an important correlation with statistical properties of natural scenes. Following this 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 principle, Gabor-like features

⋆ To whom correspondence should be addressed.

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 657–664, 2007.

resembling the receptive fields of simple cells in the primary visual cortex(V1) have been derived either by imposing sparse over-complete representations[6] or statistical independence as in Independent Component Analysis(ICA)[8].

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 begun to bring this question into consideration. Hyvarinen and Hoyer[9,10] modelled receptive fields of complex cells and Van Hateren [12]obtained spatiotemporal receptive fields of complex cells. Grimes and Rao[14] proposed a bilinear generative model to study the translation-invariance. Berkes[7] investigated temporal slowness as a learning principle for receptive fields using slow feature analysis. However, there are few models in the literatures perceiving transformations of objects or images. To investigate the problem, we apply ICA to learning from natural scenes the transformation-invariant features, and then use these features to construct a model for transformation-invariant perception. The goal of the model is to perceive transformation of patches from natural images.

The rest of the paper is organized as follows. Section 2 introduces a method for learning transformation-invariant basis functions and then propose a model for perception of transformation invariance. In section 3, we will demonstrate these basis functions and perceptual simulation results. The final section gives the comparison with other related works and models.

2 The Invariance Perception Model

In this section, we first introduce the method for learning the transformation-invariant basis functions. Then we propose a perceptual model for perception of transformation invariance.

2.1 Method for Learning Invariant Basis Functions

To obtain transformation-invariant basis functions, the training data sets should have the possession of transformational properties. The method for generating the training data will be introduced in section 3.1. Applying ICA on the training data, sequences of patches with the parameter $\alpha_i (i = 1, ..., M)$, yields transformation-invariant basis functions with parameters same as patches. For simple explanation of the method, shown in Fig.1, we use the rotation transformation and the resulting rotation basis functions.

We briefly introduce the learning algorithm of ICA for training sparse basis functions. For the standard ICA model $x = \mathbf{W}u$, Cichocki et al.[5] used the Kullback-Leibler divergence between the distribution $p(\mathbf{x}; \mathbf{W})$ of obtained by the actual value $\mathbf{W}$ and the reference distribution $q(\mathbf{x})$ to give the cost function as

$$R(\mathbf{x},\mathbf{W}) = -\frac{1}{2} \log |det(\mathbf{W}\mathbf{W}^T)| - \sum_{i=1}^{n} E \log q_i(x_i). \quad (1)$$

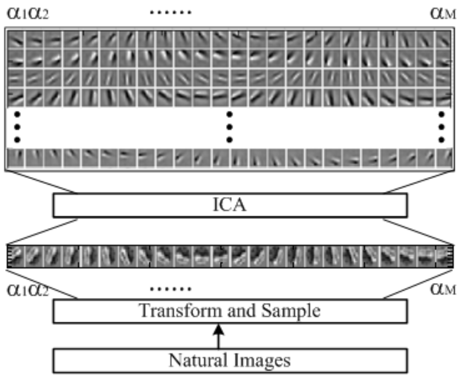

Fig. 1. Method for learning transformation-invariant basis functions. The input data is a set of natural images. Patches selected from the images are transformed and feed to the ICA algorithm.

Applying the Natural Gradient rule to the cost function, the learning algorithm of $\mathbf{W}$(the corresponding basis functions $\mathbf{A} = \mathbf{W}^{-1}$) can be described[3,4] as

$$\triangle \mathbf{W} = -\eta(t) \frac{\partial R}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W} = \eta(t)[\mathbf{I} - \langle \varphi[\mathbf{x}(k)] \mathbf{x}^T(k) \rangle] \mathbf{W}, \qquad (2)$$

where, $\varphi_i(x_i) = -\frac{q_i^{'}(x_i)}{q_i(x_i)}$. $q(x_i)$ is a supergaussian probability distribution, for instance, the Laplace pdf.

2.2 Model for Perception of Transformation Invariance

In this section, we will propose a model for transformation-invariant perception, shown in Fig. 2. The invariance perception model consists of three layers. The first layer is to receive the input patterns which are two patches with parameters of α_i and α_j, respectively. Here, α_i and α_j belong to a same group of parameters. For rotation, α is in the range of zero and three hundred sixty degree by an interval of fifteen. For scaling, α in the range of from one to two times by ten percent. And, for translation, α in the range of size of input images. For simplicity, we only discuss in detail rotational samples and basis functions in the model. The middle layer of the model is to sparsely represent input patterns with a group of basis functions which is one of three groups respectively including translational, rotational, and scaling bases, shown in Figs.{3,4,5}.

After the neurons respond to 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 $\mathbf{X}^{t_2}_{\alpha_j}(j = 1, ..., M)$, and of which the maximum is selected to determine the relative dispersion. The index (i, j) of the maximum in coefficient matrix will tell us the relative dispersion such as counter-clockwise rotation angle $\Delta\theta$, translational distance Δd, and scaling ratio

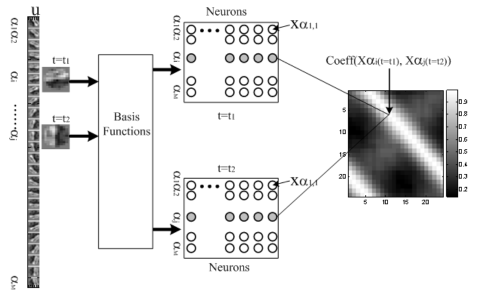

Fig. 2. Model for transformation-invariant perception. For example, the input patterns are the rotational data. $x^{t_1}_{\alpha_{i,k}}(k = 1, 2, \cdots, N)$ denotes the response of the k-th neuron in the row α_i responding to stimulus u_{α_i} at time t_1 through the basis function $\alpha_{i,k}$. And so does 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 α_i of basis functions. Namely, $\mathbf{X}^{t_1}_{\alpha_i} = [x^{t_1}_{\alpha_{i,1}}, x^{t_1}_{\alpha_{i,2}}, \cdots, x^{t_1}_{\alpha_{i,N}}]^T$.

Δr. It is necessary to note that we only need the relative transformation, not the absolute value of parameters of the stimuli. For rotation, if $j \geq i$, $\Delta\theta = (j - i) \times 360/M$; otherwise, $\Delta\theta = (M + j - i) \times 360/M$. For translation, i and j have their corresponding coordinates (x_i, y_i) and (x_j, y_j), respectively. We can calculate the relative translation distance $\Delta d = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ and the moving direction according the relative position of coordinates (x_i, y_i) and (x_j, y_j). For scaling, if $j \geq i$, $\Delta r = r_j - r_i$; otherwise, $\Delta r = r_i - r_j$. Here, r_j and r_i denote the scaling ratio of the *j-th* and *i-th* subsets of basis functions, respectively.

3 Simulations and Results

We present experimental results to verify the performance of our proposed model and the learning algorithm. First we present the basis functions of transformation invariance including translation, rotation and scaling. Then, as an example, the rotation-invariant perception is discussed.

3.1 Training Data

To learn basis functions from natural scenes, we sample a sequence of small patches of size 10×10 from a set of big natural images by three methods of transformations such as translating, rotating, and scaling. This three data sets are used to learn transformation-invariant basis functions. For example, the sampling method of rotational data set is described in detail as follows.

A sampling window is randomly located on a big natural scene and a patch is selected. 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 next one, till twenty-four times. Similarly, the total twenty-four of patches are sampled and then reshaped to one column vector as a sample, size of 2400-by-1.

We select patches from a set of big natural images by the above sampling methods and generate three data sets which are composed of 20000 samples respectively. All data sets are then low-pass filtered by reducing the dimension of the data vector by principle component analysis (PCA), retaining the 100 principal components with the largest variances, after which the data is whitened by normalizing the variances of the principal components. These preprocessing steps are essentially similar to those used in [6,9].

3.2 Transformation-Invariant Basis Functions

Respectively using the translational, scaling, and rotational training data to learn transformation-invariant basis functions, the translation-, scaling-, and rotation-invariant basis functions are yielded, shown in Figs.{3,4,5}. From these figures, we note that the Gabor-like basis functions, which are localized, oriented, and bandpass, resemble receptive fields of simple cells found in V1[13]. Meanwhile, there also are different characteristics, as follows, among three types of basis functions.

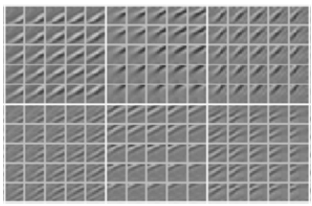

Fig. 3. Subsets of translation-invariant basis functions

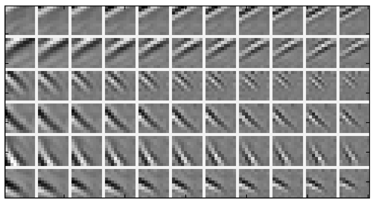

Fig. 4. Subsets of scaling-invariant basis functions

Translation-invariant basis functions. In Fig. 3, bigger rectangles composed of 5×5 basis functions with the same orientation are similar to the receptive fields of complex cells which activate while the same orientational contents are moving within their receptive fields.

Scaling-invariant basis functions. Fig. 4 shows that basis functions in one row represent the receptive field of a complex cell which performs the perception of scaling invariance. Those in one row are subsets with the same scaling. The scaling interval is 10%.

Rotation-invariant basis functions. In Fig. 5(*Left*), a group of basis functions in one 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

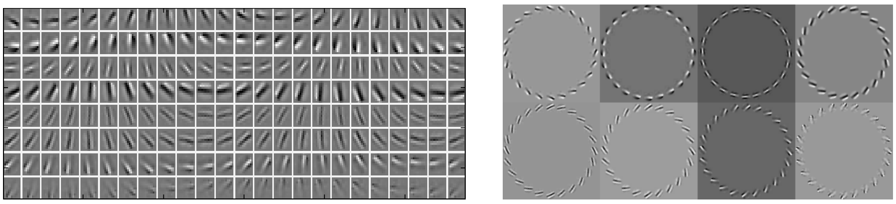

Fig. 5. Subsets of rotation-invariant basis functions. *Right*: basis functions (in the *left*) are rearranged in counter-clockwise along the circumference with an interval of fifteen degree. Every circle includes basis functions in one row (*left*).

column are a 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 Fig. 5(*Right*). Every circle resembles the receptive field of a complex cell.

3.3 Perception Experiments

For anyone of three transformations: translation, rotation, and scaling, the same experimental method is used and the invariant results are easy to obtain. Here, the invariance means that complex cells maintain their existing states, while a patch is moving, rotating and scaling within their receptive fields. In other words, we can recognize the same object however it moves, rotates, and scales within our field of vision.

An example of rotation perception is introduced and its goal is to calculate the relative rotation angle. According the perception model in section 2.2, two input patterns with different rotational angles are needed.

Randomly select two image patches u_{α_i} and u_{α_j}(i.e. i=6, j=11) from a sample data which is composed of twenty-four patches, shown in Fig. 6. For example, the sixth and eleventh of stimuli represents, respectively, rotational angles of ninety and one hundred and sixty-five in degree. The sixth patch is first input to the perception model at time $t_1$and the eleventh is second at time t_2.

Computing the responses of neurons at time t_1 and t_2 and the matrix of the correlation coefficients Coeff$(X_{\alpha_k}^{t_1}, X_{\alpha_l}^{t_2})(k, l{=}1, 2, \cdots, M)$, here M=24. Find the max value from any row in the matrix and obtain its corresponding index of its row and column, i.e. at the first row, the index of max value is (1,6). So, we know the relative rotation angle is (6-1)×15=75 degree. It is necessary to note that we only need the relative transformation, not the absolute value of angles of the stimuli.

In more detail, at time t_1 and t_2, the responses $X_{\alpha_6}^{t_1}$ and $X_{\alpha_{11}}^{t_2}$ are plotted in the last column of Fig.6. It is easy to know $X_{\alpha_6}^{t_1}$ and $X_{\alpha_{11}}^{t_2}$ are very similar to each other. The difference between $X_{\alpha_6}^{t_1}$ and $X_{\alpha_{11}}^{t_2}$, plotted in Fig. 6, shows

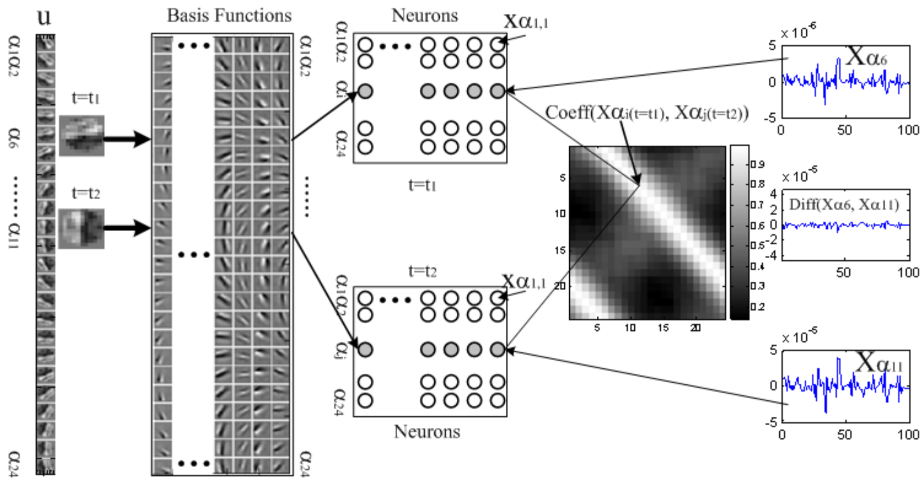

Fig. 6. Rotation-invariant perception

the rotation invariance of neuronal responses while the input pattern is rotating from time t_1 to t_2.

4 Discussions and Conclusions

We have proposed a method for learning transformation-invariant basis functions and a model for perception of transformation invariance. Computer simulation results show that our proposed model do work well in simulating the perceptual function of transformation invariance in the brain. Our proposed model has some different properties compared with others such as bilinear generative models [14] and Topo-ICA[11].

First, bilinear generative models[14] proposed by Grimes and Rao only study the translation invariance and however, ours is able to provide more transformation invariant basis functions such as translational, rotational and scaling basis functions. Our model also performs perception of the three types of transformations. On the other hand, our algorithm is much simpler whereas that of the bilinear model is more complex. Second, the Topo-ICA model[11] provided by Hyvarinen et al. considered the second-order correlation of responses of simple cells, but the Topo-ICA model cannot produce overcomplete basis functions because of constrains of orthogonality.

Our future work will focus on learning other transformational basis functions such as three dimensional geometry transformations and on transformational perception of more complex stimuli. We are also going to extend the model to a framework for learning other transformation-invariant basis functions and perception of other transformation such as view changes.

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: Rosenblith, W.A. (ed.) Sensory Communication, pp. 217–234. MIT Press, Cambridge (1961)
2. Barlow, H.B.: Redundancy reduction revisited. Network: Computation in Neural Systems 12, 241–253 (2001)
3. Zhang, L., Cichocki, A., Amari, S.: Natural Gradient Algorithm to Blind Separation of Over-determined Mixture with Additive Noises. IEEE Signal Processing Letters 6(11), 293–295 (1999)
4. Zhang, L., Cichocki, A., Amari, S.: Self-Adaptive Blind Source Separation Based on Activation function Adaptation. IEEE Transactions on Neural Networks 15(2), 233–244 (2004)
5. Cichocki, A., Zhang, L.: Two-stage bink deconvolution using state-space models. In: proceedings of the Fifth International Conference on Neural Information Processing, Kitakyushu, Japan, pp. 729–732 (1998)
6. Olshausen, B.A., Field, D.J: Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Research 37, 3311–3325 (1997)
7. Berkes, P., Wiskott, L.: Slow feature analysis yields a rich repertoire of complex cell properties. Journal of Vision 5, 579–602 (2005)
8. Bell, A.J., Sejnowski, T.J.: The independent component of natural scenes are dege filters. Vision Research 37, 3327–3338 (1997)
9. Hyvarinen, A., Hoyer, P.O.: Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. Neur. Comp. 12(7), 1705–1720 (2000)
10. Hoyer, P.O., Hyvarinen, A.: A multi-layer sparse coding network learns contour coding from natural images. Vision Research 42(12), 1593–1605 (2002)
11. Hyvarinen, A., Hoyer, P.O.: Emergence of Topography and Complex Cell Properties from Natural Images using Extensions of ICA. Advances in Neural Information Processing Systems (NIPS*99) 12, 827–833 (2000)
12. Van Hateren, J.H., Ruderman, D.L.: Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. Proc. R. Soc. London B 265, 2315–2320 (1998)
13. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology (London) 195, 215–243 (1968)
14. Grimes David, B., RAO, R.P.N.: Bilinear sparse coding for invariant vision. In: Neural computation, vol. 17(11), pp. 47–73. MIT Press, Cambridge (2005)