BLIND SOURCE SEPARATION BASED ON DUAL ADAPTIVE CONTROL

Ding Liu[1] *Xiaoyan Liu*[2] *Fucai Qian*[1] *Han Liu*[1]

[1]Xi'an University of Technology, Xi'an, China, 710048

[2] University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

E-mails：liud@xaut.edu.cn, x22liu@engmail.uwaterloo.ca, qianfc@xaut.edu.cn

ABSTRACT

This paper presents a new method for Blind Source Separation (BSS) based on dual adaptive control, which allows successful separation of linear mixtures of independent source signals. The method reformulates a BSS problem to get a dual adaptive control problem. Then a Sigmoid MLP neural network is used to approximate the widesence-mixing matrix defined in the BSS problem. By solving the dual adaptive control problem, in which unknown parameters of the neural network are estimated by applying the Extended Kalman Filter, we then obtain the widesense-mixing matrix. Experimental results show that individual source signals can be separated effectively from the known linear mixture signals using this method. And faster convergence speed as well as good performance can be achieved.

Key words Blind Source Separation (BSS), Dual Adaptive Control, Extended Kalman Filter (EKF)

1.INTRODUCTION

Blind Source Separation (BSS) has potential applications in many areas such as pattern recognition, communication, and medical image processing [1]. Consider m observed signals, which are linear mixtures of n independent source signals that are unknown. We assume $m = n$ in this paper. The BSS problems are to extract individual unknown source signals from these known mixtures, with the prerequisite that the source signals are statistically independent and non-Gaussian distributed (or at most one is Gaussian distributed) [8].

The general approach used in BSS problems is to find a linear transformation matrix or separation matrix W in order to extract individual source signals from the known mixtures. A number of researchers have investigated various methods to find a separation matrix W [2][4][5][6]. Bell and Sejnowski (1995) presented an information maximization (Informax) algorithm using a feedforward neural network, which can implement separation of mixtures of a number of supergaussian sources [2]. T. W. Lee et al. (1999) extended the Informax algorithm based on natural gradient (Amari) [4][6], which can effectively separate the mixtures of subgaussian and supergaussian sources. Nandi et al. (2000) developed an adaptive method to estimate an orthogonal rotation matrix that implemented successful separation. This algorithm involves a great deal of complex number calculation, which affects the calculation efficiency [1]. The method presented in this paper reconstruct BSS problem into a dual adaptive control problem based on their essential similarity. A Sigmoid MLP (multilayer perceptron）neural network is applied in the control problem to approximate the widesense-mixing matrix in BSS problems, which is defined in section 2. In section 3 and 4 the dual adaptive control problem is solved and the unknown network parameters are estimated using extended Kalman filter (EKF) to obtain a good separation matrix W. Simulation experiments show that this method can separate the mixture signals effectively and the algorithm converges rapidly.

2.BLIND SOURCE SEPARATION (BSS)

Consider m observed signals

$$X(k) = [x_1(k), \cdots, x_m(k)]^T$$

which are linear mixtures of n statistically independent source signals $S(k) = [s_1(k), \cdots, s_n(k)]^T$, which are unknown. Assume that there exists no time delay when mixing these sources. Then we can write the BSS equation in the form

$$X(k) = MS(k) + e(k) \quad (2\text{-}1)$$

where M denotes the linear mixing matrix and $rank(M) = n$, $e(k)$ is n-dimensional observation noise vector with covariance matrix $\sigma^2 I$. If not

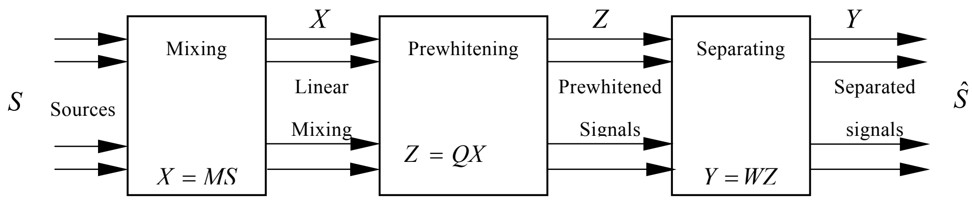

Fig. 1. BSS Block Diagram

considering the effect of noise, (2-1) is written as

$$X(k) = MS(k) \tag{2-2}$$

The above block diagram illustrates the process to solve BSS problems. Block 1 represents the linearly mixing process of source signals. We know nothing about the linear mixing matrix M except that the sources are independent and non-Gaussian distributed. Since the mixtures are not equally distributed in all directions, they should be first prewhitened as shown in block 2. Let the prewhitening matrix be Q and we thus get the prewhitened Gaussian signals $Z(k) = QX(k) = QMS(k)$, which are independent. Then we define a widesense-mixing matrix A satisfying $A = QM$, i.e. the product of the prewhitening matrix and the mixing matrix. Hence, as shown in Block 3, prewihitened signals $Z(k)$ are separated by solving the equation

$$Y(k) = WZ(k) = WAS(k) = \hat{S}(k) \tag{2-3}$$

where $Y(k)$ denotes the separated signals or the source signal estimations $\hat{S}(k)$. It can be seen from (2-3) that if the separation matrix W equals the inverse of the widesense-mixing matrix, i.e. $W = A^{-1}$, then we have $Y(k) = S(k)$, that is to say, the sources are successfully extracted from the mixed signals.

3.DUAL ADAPTIVE CONTROL FOR STOCHASTIC SYSTEMS

In 1960s, Russian scholar Feldbaum introduced an important dual property for the stochastic controller: The optimal control has dual effects of both learning and regulating upon the stochastic system. For learning, the control introduces sufficient stimuli to the system in order to obtain abundant information that favors parameter identification; for regulating, it always strives to make actual system output follow the expected output value, i.e. to stabilize it at the expected level. These two properties usually conflict with each other for the identification of unknown parameters needs relatively large actuation signals to detect the system status while regulation of the system needs only minor signals [9][10]. Dual control strategy achieves optimal tradeoff between learning and regulating so that the overall control performance of the system can be improved [7].

The objective of dual control is to find a control input sequence $u(k)$, which minimizes the following N-step minimum variance performance index

$$J_{dual} = E\{\sum_{k=0}^{N-1}[y(k+1) - y_r(k+1)]^2 \mid Y^k\} \tag{3-1}$$

where $y_r(k)$ is the desired system output and $y(k)$ the actual system output. Y^k represents the total information in instant k, defined as $Y^k = \{y(k),\ldots,y(0),u(k-1),\ldots,u(0)\}$ and $E\{\cdot\}$ the conditional expectation with respect to all the random system variables given that Y^k. Theoretically, the optimal control can be achieved by solving dynamic programming, i.e. Bellman equations. But in most cases Bellman equations are high-dimensional stochastic nonlinear equations and thus very difficult to get the optimal solution. In this paper a Sigmoid MLP neural network is used to solve the control problem and the extended Kalman filter (EKF) is applied to estimate the unknown network parameters [7].

4.BSS SOLUTIONS BASED ON DUAL ADAPTIVE CONTROL METHOD

In this section a new method for BSS problems based on dual adaptive control is derived. In order to formulate a proper control system model, first we consider m observed signals or mixture signals $X(k)$, the only known information for the BSS problem. Since $X(k)$ are not equally distributed in all directions, we prewritten it to get independent Gaussian signals $Z(k)$ with zero mean value and unity covariance matrix [1]. And as shown in Fig.1, $Z(k)$ satisfies $Z(k) = QX(k) = QMS(k) = AS(k)$. Hence a Sigmoid MLP neural network is used to establish a nonlinear function matrix to approximate the widesense-mixing matrix A satisfying $A = QM$. According to (2-1), we get a dual adaptive control problem written in the form

$$y(k) = f(\cdot)u(k) + e(k) \qquad (4\text{-}1)$$

where m-dimensional vector $y(k)$ is the system output and n-dimensional vector $u(k)$ the system control, and we assume $m = n$. $e(k)$ denotes n-dimensional measurement noise vector with zero mean value and covariance matrix $\sigma^2 I$. $f(\cdot)$ represents the unknown nonlinear network function matrix that is used to approximate unknown widesense-mixing matrix A. According to (4-1), we see that system output $y(k)$ is the measurement of n linear mixture signals, which means that the desired system output should be the prewhitened signals $Z(k)$. And the control $u(k)$ is equivalent to the estimation of n unknown sources $\hat{S}(k)$. Without loss of generality, we assume that $f(\cdot)$ is a function of $Z(k)$, i.e. $f(Z(k))$. Then the system equation (4-1) is written as

$$y(k) = f(Z(k))u(k) + e(k) \qquad (4\text{-}2)$$

where $f(Z(k)) = \begin{bmatrix} f_{11}(Z(k)) & \cdots & f_{1n}(Z(k)) \\ \vdots & \ddots & \vdots \\ f_{n1}(Z(k)) & \cdots & f_{nn}(Z(k)) \end{bmatrix}$.

The Sigmoid MLP neural network applied includes a hidden layer with $n \times n$ nodes and one output node whose output can be defined as

$$f_{ij}(Z(k)) = c_{ij}(k)\Phi_{ij}(k), \quad i, j = 1, \cdots, n \qquad (4\text{-}3)$$

where c_{ij} denotes weighted coefficient of output layer and Φ_{ij} the Sigmoid activation function with the following form

$$\Phi_{ij}(k) = \frac{1}{1 + e^{-w_{ij}^T(k) \cdot Z_a(k)}} \qquad (4\text{-}4)$$

where w_{ij} is the weighted coefficient vector of the $i \times j$ th neuron and $Z_a(k) = [Z^T(k), 1]^T$ is $(n+1)$-dimensional augmented state vector of the system in which constant 1 is used as the bias input of the neuron.

In order to estimate the unknown network parameters, we construct an $(n(n+2) \times n)$-dimensional parameter matrix $w(k)$ as follows

$$w(k) = \begin{bmatrix} c_{11} & \cdots & c_{n1}, w_{11} & \cdots & w_{n1} \\ \vdots & \ddots & \vdots \quad \vdots & \ddots & \vdots \\ c_{1n} & \cdots & c_{nn}, w_{1n} & \cdots & w_{nn} \end{bmatrix}^T \qquad (4\text{-}5)$$

Assume that there exists an estimated $w^*(k)$ to make the difference between the approximation of nonlinear function matrix $f(Z(k))$ and the widesense-mixing matrix A small enough:

$$w^*(k) = \begin{bmatrix} c_{11}^* & \cdots & c_{n1}^*, w_{11}^* & \cdots & w_{n1}^* \\ \vdots & \ddots & \vdots \quad \vdots & \ddots & \vdots \\ c_{1n}^* & \cdots & c_{nn}^*, w_{1n}^* & \cdots & w_{nn}^* \end{bmatrix}^T$$

Then based on (4-2) and (4-3), the state-space model of the problem is written as

$$w^*(k+1) = w^*(k) \qquad (4\text{-}6a)$$

$$y(k) = h(w^*(k), Z(k), u(k)) + e(k) \qquad (4\text{-}6b)$$

which represent the state and the output equations respectively. And in the model we define that

$$h(w^*(k), Z(k), u(k)) = f(Z(k))u(k) = \begin{bmatrix} c_{11}\Phi_{11} & \cdots & c_{1n}\Phi_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1}\Phi_{n1} & \cdots & c_{n1}\Phi_{n1} \end{bmatrix} u(k) \qquad (4\text{-}7)$$

which is a nonlinear function of $w^*(k)$. Since the model is nonlinear, we use the extended Kalman filter (EKF) to estimate the unknown parameter matrix and the estimation equations are [7]

$$K(k) = P(k)\tilde{\nabla}_h^T(k)[\sigma^2 + \tilde{\nabla}_h(k)P(k)\tilde{\nabla}_h^T(k)]^{-1} \qquad (4\text{-}8a)$$

$$P(k+1) = [I - K(k)\tilde{\nabla}_h(k)]P(k) \qquad (4\text{-}8b)$$

$$w(k+1) = w(k) + K(k)[Z(k) - h(w^*(k), Z(k), u(k)]^T \qquad (4\text{-}8c)$$

where $n(n+2)$-dimensional matrix $P(k)$ denotes the covariance matrix of parameter estimation error, $n(n+2)$-dimensional vector $K(k)$ filter gain matrix. $\tilde{\nabla}_h(k)$, an important variable in this method, is defined as the function of $u(k)$ in the form that

$$\tilde{\nabla}_h(k) = u^T(k)\tilde{\nabla}_f(k) \qquad (4\text{-}9)$$

where

$$\tilde{\nabla}_f(k) = \begin{bmatrix} \Phi_{11}(k) & \cdots & \Phi_{n1}(k), & \Psi_{11}(k) & \cdots & \Psi_{n1}(k) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \Phi_{1n}(k) & \cdots & \Phi_{nn}(k), & \Psi_{1n}(k) & \cdots & \Psi_{nn}(k) \end{bmatrix}$$

denotes the transpose of the gradient matrix of $f(Z(k))$ with respect to $w^*(k)$ given $w^*(k) = \hat{w}(k)$ and $\hat{w}(k)$ is the parameter estimation at interval k. The entries of $\tilde{\nabla}_f(k)$ is obtained by the following equations:

$$\Phi_{ij}(k) = \frac{\partial f_{ij}(k)}{\partial c_{ij}} = \frac{\partial c_{ij}(k)\Phi_{ij}(k)}{\partial c_{ij}} = \frac{1}{1 + e^{-\hat{w}_{ij}^T(k) \cdot Z_a(k)}} \qquad (4\text{-}10a)$$

$$\Psi_{ij}(k)=\frac{\partial f_{ij}(k)}{\partial c_{ij}}=\frac{\partial c_{ij}(k)\Phi_{ij}(k)}{\partial w_{ij}}$$

$$=\hat{c}_{ij}(k)\cdot e^{-\hat{w}_{ij}^T(k)\cdot Z_a(k)}\cdot\Phi_{ij}^2\cdot Z_a^T(k)$$

$$i,j=1,2,\cdots,n. \quad \text{(4-10b)}$$

Assume that the initial parameter vector $w^*(0)$ is Gaussian distributed with mean value of m_0 and covariance matrix R_0 and $w^*(k)$ is has conditional Gaussian distribution with mean value $\hat{w}(k)$ and covariance matrix $P(k)$ given $Z(k)$. And the Gaussian distribution here is approximated and different from the linear Kalman filter. For the output $y(k)$, based on the nonlinear relation described in (4-6b), we calculate the first order Taylor series with respect to $\hat{w}(k)$ in order to obtain the conditional distribution of $y(k)$ given $Z(k)$ and then have

$$y(k)\approx h(\hat{w}(k),Z(k),u(k))+\tilde{\nabla}_h(k)[w^*(k)-\hat{w}(k)]+e(k) \quad \text{(4-11)}$$

Since $w^*(k)$ has approximate conditional Gaussian distribution given $Z(k)$, based on (4-11), $y(k)$ also has approximate conditional Gaussian distribution given $Z(k)$ [7], whose mean value is $h(\hat{w}(k),Z(k),u(k))$ and covariance matrix is $\tilde{\nabla}_h(k)P(k)\tilde{\nabla}_h^T(k)+\sigma^2 I$.

Consider the performance index used in dual adaptive control problem [7]

$$J(k)=E\{[y(k)-Z(k)]^T[y(k)-Z(k)] + qu^T(k)u(k)-\lambda v^T(k)v(k)\,|\,Z(k)\} \quad \text{(4-12)}$$

where $E(\cdot\,|\,Z(k))$ represents the conditional expectation given $Z(k)$ and the weighted coefficient satisfies $q\geq 0$, $0\leq\lambda\leq 1$. $v(k)$ denotes system innovations that has the form

$$v(k)=y(k)-h(w^*(k),Z(k),u(k))$$

In order to minimize the performance index, we first transform it into an explicit function of the system control $u(k)$. For convenience, represent $h(w^*(k),Z(k),u(k))$ as $h(k)$.

$$J(k)=E\{[y(k)-h(k)+h(k)-Z(k)]^T[y(k)-h(k) + h(k)-Z(k)]+qu^T(k)u(k)-\lambda[y(k)-h(k)]^T [y(k)-h(k)]\,|\,Z(k)\}$$

$$=E\{(1-\lambda)[y(k)-h(k)]^T[y(k)-h(k)]+[h(k)-Z(k)]^T [h(k)-Z(k)]+qu^T(k)u(k)\,|\,Z(k)\}$$

$$=(1-\lambda)[\tilde{\nabla}_h(k)P(k)\tilde{\nabla}_h^T(k)+\sigma^2]+qu^T(k)u(k) +[h(k)-Z(k)]^T[h(k)-Z(k)] \quad \text{(4-10b)}$$

Then based on (4-7) and (4-9), reduce the equation into

$$J(k)=(1-\lambda)[u^T(k)\tilde{\nabla}_f(k)P(k)\tilde{\nabla}_f^T(k)u(k)] +(1-\lambda)\sigma^2+qu^T(k)u(k)+[f(Z(k))u(k)-Z(k)]^T [f(Z(k))u(k)-Z(k)]$$

$$=u^T(k)[(1-\lambda)\tilde{\nabla}_f(k)P(k)\tilde{\nabla}_f^T(k)+qI +f^T(Z(k))f(Z(k))]u(k)-2u^T(k)\cdot f^T(Z(k))\cdot Z(k) +Z^T(k)Z(k)+(1-\lambda)\sigma^2 \quad \text{(4-13)}$$

Let the derivative of the reduced performance index with respect to $u(k)$ be zero, i.e.

$$\frac{\partial J(k)}{\partial u(k)}=2[(1-\lambda)\tilde{\nabla}_f(k)P(k)\tilde{\nabla}_f^T(k)+qI +f^T(Z(k))\cdot f(Z(k))]u(k)-2f^T(Z(k))\cdot Z(k)=0$$

Solve the equation and get the optimal system control as follows

$$u^*(k)=[(1-\lambda)\tilde{\nabla}_f(k)P(k)\tilde{\nabla}_f^T(k)+ f^T(Z(k))\cdot f(Z(k))+qI]^{-1}f^T(Z(k))\cdot Z(k) \quad \text{(4-14)}$$

Substitute (4-14) into the EKF equations (4-8a), (4-8b) and (4-8c) to estimate the unknown parameter matrix so that we can get the estimation of the widesense-mixing matrix A, i.e. $A^*=f^*(Z(k))$. As a result, we then can extract each source from the mixture signals by solving the following equation

$$\hat{S}(k)=[f^*(Z(k))]^{-1}\cdot Z(k) \quad \text{(4-15)}$$

where $[f^*(Z(k))]^{-1}$ is the separation matrix for the BSS problem.

5.EXPERIMENTAL RESULTS

5.1.Two source signals separation

Collect two sound signals "good morning" and "KongBangWa" which means "good morning" in Japanese as the sources (sampling frequency is 16kHz). Randomly choose the entries of the mixing matrix M over the range (0, 1) and we have $M=\begin{bmatrix}0.6575, & 0.3667\\ 0.4548, & 0.4751\end{bmatrix}$. Use the mixing matrix and get the linear mixture signals $X=[x_1,x_2]$. In the simulation experiment 2.5×10^4 points are sampled and only 50 of the total sampling points are calculated. Thus the learning of the nonlinear function matrix $f(Z(k))$ converges rapidly. Then use (4-15) we get the two separated sound signals.

First we provide the source signals which are actually unknown below.

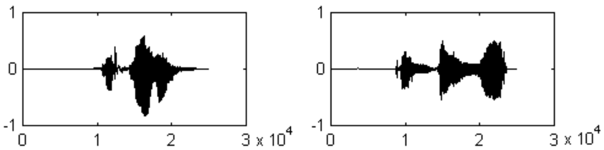

Fig. 2a. Diagram of two source signals

Then the diagram of the mixed sound signals (without noise) is given in Fig.2b.

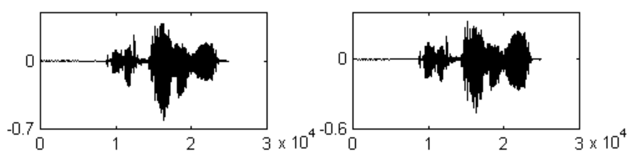

Fig. 2b. Diagram of two mixed signals

Finally the two separated sound signals are plot as follows:

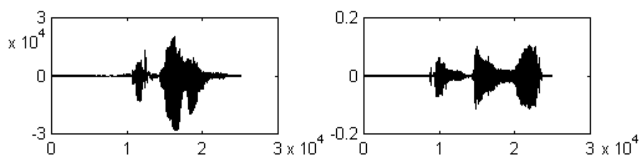

Fig. 2c. Diagram of two separated signals

Simulation results show good separation performance of the method. And the two separated sound signals can be clearly played.

5.2. Multi-source signals separation

In this experiment, we add a third sound signal "GongNengWanMei, AoMiaoWuQiong" into the sources (sampling frequency is still 16kHz) and the experimental results are shown below.

First the three source signals are plot as shown in Fig. 3a.

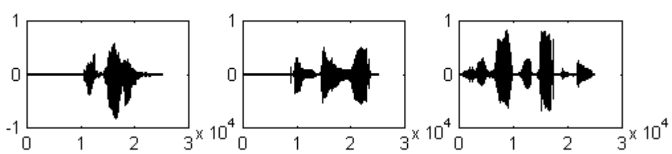

Fig. 3a. Diagram of source signals

Similar to experiment A, we randomly choose a mixing matrix whose entries are over the range (0, 1) and mix the sources linearly without noise perturbation. See the mixed signals in Fig. 3b.

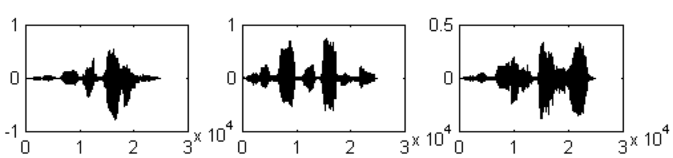

Fig. 3b. Diagram of mixed signals

Finally the separated sound signals are obtained, as shown in Fig. 3c.

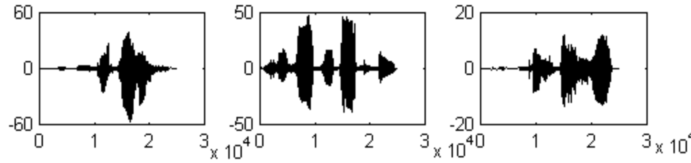

Fig. 3c. Diagram of separated signals

It can be seen that the method is effective to separate multi-source signals. But because of the limitation of dual adaptive control in learning performance and of the calculation errors, a lot more work need to be done to improve the separation performance in the case of more source signals.

6.CONCLUSIONS

A new method to solve BSS problems based on dual adaptive control strategy has been presented in this paper. The problem is reformulated to establish a typical dual adaptive multivariable control problem. To solve the control problem, a Sigmoid MLP neural network is used and the extended Kalman filter is also applied in order to get accurate parameter estimation. Then the widesense-mixing matrix of the BSS problem is approximated and thus the separation matrix can be obtained. This method doesn't have to calculate all the sampling points of the signals and thus greatly improve the computation efficiency. Simulation results illustrate that the method can successful separate two source signals from the linear mixture and converge very rapidly. Further research is needed to improve the method in the case of multi-source signals. A better performance index for the BSS problem is expected, which may improve the separation performance greatly.

7.REFERENCES

[1] V. Zarzoso and A. K. Nandi, "Adaptive Blind Source Separation for Virtually Any Source Probability Density Function," *IEEE Trans. Signal Processing*, vol. 48, pp. 477-487, Feb. 2000.

[2] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, 1995.

[3] Han Liu, Ding Liu, Xiaoyan Liu, "Blind Source Separation Using a Rotation Matrix Identification Algorithm," *Proceedings of International Conference of IEEE on Info-tech and Info-net* (IEEE Catalog Number: 01EX479), ICII, Beijing, China, Oct. 29-Nov. 1, 2001, pp. 232-236.

[4] S. Amari, "Natural Gradient Works efficiently in Learning," *Neural Computation*, vol. 10, pp. 251-276, 1998.

[5] T. W. Lee, M. Girolami and T. J. Sejnowski, “Independent Component Analysis Using an Extended Informax Algorithm for mixed Subgaussian and Supergaussian Sources,” *Neural Computation*, vol. 11, pp. 417-441, 1999.
[6] S. Amari, A. Cichocki and H. H. Yang, "A New Learning Algorithm for Blind Signal Separation," in NIPS'95, *MIT Press*, Cambridge, Mass., vol. 8, pp. 757-763, 1996.
[7] Simon Fabri and V. Kadirkamanathan, “Dual Adaptive Control of Nonlinear Stochastic Systems using Neural Networks,” *Automatica*, Vol. 34, No. 2, pp. 245-253, 1998.
[8] T. Wachtler and T.W. Lee, “Chromatic Structure of Natural Scenes,” *Optical Society of America*, Vol. 18, No. 1, pp. 65-77, Jan. 2001.
[9] A. A. Feldbaum, “Dual Control Theory Ⅰ-Ⅱ,” *Automat. Remote Contr.*, vol. 21, pp. 874-880, 1032-1039, 1960.
[10] A. A. Feldbaum, “Dual Control Theory Ⅲ-Ⅳ,” *Automat. Remote Contr*, vol. 22, pp. 1-12, 109-121, 1961.