

## Slow Feature Extraction Algorithm Based on Visual Selection Consistency Continuity and Its Application



Hong Yang, Yanming Zhao\*, Guoan Su, Xiuyun Liu, Songwen Jin, Haoyang Fan, Yuhui Shang

School of Mathematics and Computer Science, Hebei Normal University for Nationalities, Chengde 067000, China

Corresponding Author Email: zhaoyanming008@163.com

https://doi.org/10.18280/ts.380604	ABSTRACT	

Received: 5 September 2021 Accepted: 18 November 2021

#### Keywords:

visual invariance, visual selection consistency continuity, natural image, slow feature, Lipschitz consistency The conventional slow feature analysis (SFA) algorithm has no support of computational theory of vision for primates, nor does it have the ability to learn the global features with visual selection consistency continuity. And what is more, the algorithm is highly complex. Based on this, Slow Feature Extraction Algorithm Based on Visual selection consistency continuity and Its Application was proposed. Inspired by the visual selection consistency continuity theory for primates, this paper replaced the principal component analysis (PCA) method of the conventional SFA algorithm with the myTICA method, extracted the Gabor basis functions of natural images, initialized the basis function family; it used the feature basis expansion algorithm based on visual selection consistency continuity (the VSCC\_FBEA algorithm) to replace the polynomial expansion method in the original SFA algorithm to generates the Gabor basis functions of features with long and short-term visual selectivity in the family of basis functions, which solved the drawbacks of the polynomial prediction algorithm; it also designed the Lipschitz consistency constraint, and proposed the Lipschitz-Orthogonal-Pruning-Method (LOPM algorithm) to optimize the basis function family into an over-complete family of basis functions. In addition, this paper used the feature expression method based on visual invariance theory (visual invariance theory -FEM) to establish the set of features of natural images with visual selection consistency continuity. Subsequently, it adopted three error evaluation methods and mySFA classification method to evaluate the proposed algorithm. According to the experimental results, the proposed algorithm showed good prediction performance with respect to the LSVRC2012 data set; compared with the SFA, GSFA, TICA, myICA and mySFA algorithms, the proposed algorithm is correct and feasible; when the classification threshold of the algorithm was set at 8.0, the recognition rate of the proposed algorithm reached 99.66%, and neither of the false recognition rate and the false rejection rate was higher than 0.33%. The proposed algorithm has good performance in prediction and classification, and also shows good anti-noise capacity under limited noise conditions.

#### **1. INTRODUCTION**

In the field of image analysis and understanding, compared with fast signals, slow signals are the essential representation of invariants. They are the invariant features extracted from the high-level expression of the input high-frequency visual signals. Based on the idea of invariant learning, the slow feature analysis theory and the analysis algorithm were proposed. The algorithm is aimed to explore how to extract the slow attributes contained in high-frequency signals and the slow topology between these attributes, and has been successfully applied.

On the basis of previous studies, in 1989, Hinton [1] first proposed the basic concepts, principles and assumptions that later formed the slow feature analysis theory, and established the research foundation for slow feature analysis methods. In 2002, Wiskott and Sejnowski [2] proposed the slow feature analysis (SFA) algorithm, which conducts non-linear expansion of the feature space in an unsupervised learning manner. This algorithm effectively solved the problems of the supervised learning algorithm like insufficient sample size and constant dimensions. By this time, the theory and algorithm of slow feature analysis had been initially established, and applied research had been carried out.

In order to solve the problems in the studies on the complex cell receptive field learning models, in 2005, Berkes [3] and Wiskott applied the invariant learning results to this field and achieved great results. In 2007, Franzius et al. [4] studied the invariant relationship between hippocampal pyramidal cells and stimulation signals and revealed the stable response between the two, forming the physiological basis for the invariant learning theory.

In order to study the invariant features of the position and rotation angles of cartoon fish, in 2008, Franzius et al. [5] applied the SFA algorithm for the first time to realize invariant feature extraction and analysis, and also achieved the classification of cartoon fish based on such features. Therefore, the classification capacity of the SFA algorithm has been confirmed, laying a foundation for the theory and application of the SFA algorithm in invariant feature extraction, analysis and recognition. In order to verify the feasibility and effectiveness of the SFA algorithm in feature extraction, Kalmpfl and Maass [6] used the SFA algorithm to extract the slow features of the training data generated by the Markov

chain in 2009. The test results show that: when  $\alpha \rightarrow o$ , the slow features extracted by the SFA algorithm are equivalent to those extracted by the Fisher's Linear Discriminant FLD algorithm. In terms of feature extraction, the SFA algorithm proved to be effective and feasible, and it was also demonstrated that the training data generated by the Markov chain method have slow features.

In order to achieve the expansion of the basis function space in blind signal analysis and solve the problem of highdimensional operation, Ma et al. [7] proposed a kernel-based slow-variation feature analysis algorithm in 2011, which achieved good performance in blind signal separation and further expanded the application field of standard SFA algorithms.

The complexity of the SFA algorithm mainly stems from the slow feature generation method, which is also a problem that scholars are generally concerned about. Inspired by the divide-and-conquer strategy, Luciw et al. [8] proposed a hierarchical slow feature analysis algorithm in 2012. The algorithm made a tradeoff between the information loss generated by the hierarchy and the complexity of the algorithm, that is, it reduced computational complexity of SFA at the cost of information loss. At the same time, the basis function family also lost its over-completeness and global features.

In order to improve the problem of information loss, Escalante-B and Wiskott [9] proposed a slow feature analysis algorithm based on graph theory (GFSA) in 2013. The research proposed the idea of using the complex structure of the training graphs to train the SFA algorithm, enabling the SFA algorithm to learn the invariance of the graph structures in the training data set and realizing the extraction of slow features and slow structures of the training set. It confirmed the dual connotations of slow features.

In 2015, Zhao et al. [10] proposed a feature extraction algorithm based on the complex visual information of natural images, which uses the custom myTICA algorithm to extract the slow features and the slow topological structures in the visual space of natural images based on visual selectivity, confirming the visual selectivity of primates determines the slow features and slow topological structures contained in natural images. In 2019, Zhao [11] bstudied an improved SFA algorithm based on visual invariance, and applied it to extract features of natural images, which achieved good results. The study confirmed the invariance of slow features and slow structures in the visual space, and the proposed algorithm proved to have good anti-noise capacity [12]. In 2020, Zhao [13] studied an improved SFA algorithm based on visual selectivity, and applied it in the feature extraction of limited defocused image sequences, which achieved good results. The study proved that in the visual space, the slow features and slow structures contained in the limited defocused image sequences are invariant, and also established an invariant feature forest.

In 2020, Zhao [14, 15] proposed an improved multi-layer LSTM algorithm based on spatial-temporal correlation, and applied it in the dynamic prediction of PM25, which achieved good results. The study proved that: in time series analysis, the improved LSTM algorithm integrating spatial features has better long-term and short-term prediction capabilities, and can realize the spatio-temporal fusion of the prediction results, which solved the problem that the traditional LSTM algorithm could not achieve the fusion of spatial features.

In recent years, the SFA algorithms have been well applied in various fields, such as human behavior recognition [16-18], blind signal analysis [7, 19], dynamic monitoring [20], 3D feature extraction [21] and multi-person path planning [22].

In summary, there has been great progress in the research on the traditional slow feature analysis (SFA) algorithm. However, it still has a number of drawbacks: 1. PCA is adopted as the feature extraction method, which ignores the visual independence between basis functions. 2. the polynomial expansion method is adopted in the expansion of basis functions, lacking the support of the computational theory of vision for primates, and making the algorithm unable to learn global features with visual selection consistency continuity, and what is more, the algorithm is highly complex. 3. The feature expression is not diversified and the slow structure of slowly-varying information is ignored.

Therefore, Slow Feature Extraction Algorithm Based on Visual selection consistency continuity and Its Application was proposed. This paper solved the drawbacks of the traditional SFA algorithm through four innovations: 1. it replaced the PCA method of the traditional SFA algorithm with the myTICA method, extracted the Gabor basis functions showing the invariance of natural images, and generated the basis function family; 2. it used the VSCC FBEA algorithm to realize the basis function expansion of features with visual selection consistency continuity and generated an overcomplete basis function family; 3. it proposed using the LOPM algorithm to optimize the over-complete basis function family, and generated the optimized over-complete basis function family; 4. It adopted the visual invariance theory -FEM to generate the set of invariable features containing the visual information of natural images and its spatial topological relations.

#### 2. THEORY OF VISUAL SELECTION CONSISTENCY CONTINUITY AND IMPROVEMENTS MADE TO THE SFA ALGORITHM

The visual selectivity theory shows that: a finite number of visual neurons in primates can express the infinite nature, which can be expressed by the Gabor basis function family. The theory of visual selectivity consists of two parts: visual selectivity and visual selection consistency continuity. Visual selectivity theory means that, in the visual space, the Gabor basis function family follows the visual selectivity rule of "similar when being close, and different when being distant"; and the visual selection consistency continuity theory means the basis function parameters in the same family are subject to the constraint of consistency continuity. To further illustrate the visual selectivity theory, one Gabor function family containing similar Gabor basis functions represent the regions with similar textures in an image; different Gabor function families represent the regions with different textures in the image; and in the same image, Gabor function families and the basis functions in each family have stable visual information and spatial topological structure of information. The slow visual feature theory shows that when the visual system of primates observes the external environment, the observation signals have slowly varying features under the cover of the rapid change process; that is, the slow variances in the Gabor basis functions, the Gabor basis function families and their visual spatial topological structures contained in the signals. The visual processing of primates is a process with parallel processing of multiple links and hierarchical serial processing in each link [23-25], where the links slowly form stable

structures and functions during the infinite image training process. Therefore, the link structure and its functions vary slowly, and the Gabor basis functions, Gabor basis function families and their visual spatial topological structures learned through links also vary slowly. The LSTM algorithm is an improved recurrent network algorithm that effectively learns or predicts the long-term and short-term memory features of time series through its own stable link structure. It can ensure prediction based on the fusion of global information and local information, and realize the visual selection consistency continuity under multi-link parallel processing. In particular, the deep LSTM algorithm can realize multi-parameter fusion prediction.

Based on the above research, this paper integrated the visual selectivity theory with the deep LSTM algorithm. It adopted the deep LSTM algorithm to predict the four key parameters of the Gabor basis functions, and generated the Gabor basis function family; inspired by the visual selection consistency continuity theory, it proposed the LOPM algorithm to optimize the generated Gabor basis function into an over-complete basis function family. It also used the VSCC\_FBEA algorithm to replace the basis function prediction method based on polynomial product in the SFA algorithm.

#### **3. PROPOSED ALGORITHM**

#### 3.1 Analysis on the existing studies

The above sections have outlined the physiological basis for and the basic theories and implementation of the SFA algorithm, and analyzed the main reasons behind the problems of the traditional SFA algorithm. Such information is summarized in the Table 1:

<b>Table I.</b> Existing research results of the SFA algorith
---

Research content	Key literatures	Feature extraction method	Feature extraction method Feature expansion method			
	Wiskott and Sejnowski [2] in 2002	PCA	Polynomial expansion	Complete set of slow features with over completeness		
	Luciw et al. [8] in 2012	PCA	Polynomial expansion	Set of slow features in hierarchical sampling, with no completeness		
Slow	Escalante-B and Wiskott [9] in 2013	PCA	Polynomial expansion	Set of graph structure features, revealing the topological stability		
features	Zhao et al. [10] in 2015	MyTICA	Based on feature probability distribution	Set of invariant features of visual spatial topological structures		
	Zhao [11, 13] in 2010-2020	MyTICA	Visual invariance and visual selectivity	Three-dimensional visual information and visual spatial topological structure information		
Improved LSTM	Zhao [15] in 2020	Long-term and short-term microscopic memory features based on spatial-temporal correlation	N/A	Global and local spatial correlation matrix (on the micro-level) with long- and short-term features		
prediction algorithm	Zhao [14] in 2021	Long-term and short-term macroscopic memory features based on spatial-temporal correlation	N/A	Global and local spatial correlation matrix (on the macro-level) with long- and short-term features		

The comparative study in Table 1 shows that the feature extraction method is gradually approaching the way the human brain understands the world: from PCA to MyTICA, this change achieved feature independence and the learning of spatial topological structures; the transformation in the feature expansion method, from polynomial expansion to the one based on feature probability distribution and then to the basis function expansion method based on visual invariance and visual selectivity, shows that the brain computing method has both theoretical and practical significance in the SFA basis function expansion algorithm.; and in terms of feature structure, from full features to hierarchical features and then to 3D visual information and visual spatial topological structure information, the expression of feature structure gradually approached the cognitive structure of the human brain.

Therefore, this paper made innovations in the extraction and expansion of slow feature basis functions and the generation and expression of slow spatial structures. It integrated the visual selectivity theory, the Lipschitz consistency continuity theory and the LSTM theory, proposed the LSTM algorithm based on visual selectivity as the basis function space expansion algorithm and also the LOPM algorithm to realize the effective fusion of the slow feature basis function expansion method and the visual selection consistency continuity, and predicted and extracted an over-complete basis set from natural images; and finally, it achieved the recognition of natural images. The innovations of the proposed algorithm are demonstrated in the following section one by one.

#### 3.2 Improvements made in the proposed algorithm

### Improvement 1: Receptive field model based on visual selectivity

The Gabor function can model the visual receptive field of visual cells. The four key parameters of the Gabor function realize the visual selectivity of primates' visual cells, but with different contributions in the selective expression. In terms of contributions, the four key parameters in the descending order are frequency, orientation, phase and visual spatial position. Therefore, based on the visual selectivity theory, the Gabor model is proposed as follows:

$$\begin{cases} gabor = gabor(\lambda_f, \lambda_o, \lambda_p, \lambda_d) \\ \lambda_f + \lambda_o + \lambda_p + \lambda_d = 1 \\ \lambda_f \ge \lambda_o \ge \lambda_p \ge \lambda_d \end{cases}$$
(1)

where,  $\lambda_f$  and  $\lambda_o$  respectively represent the visual spatial frequency and visual spatial orientation of the receptive field

of visual cells, which play a decisive role in the visual spatial distribution process- mainly determining the type and function of the receptive field of visual cells.  $\lambda_p$  and  $\lambda_d$  respectively represent the visual space hysteresis of the visual cell receptive field and the size of the distribution area of the visual cell receptive field, which play a secondary role in the visual spatial distribution process - determining the hysteresis and positions of the intra-class visual cells.

Inspired by the theory of visual selectivity, considering the different roles (contributions) of the four parameters in the Gabor function, the normalized parametric method based on contribution is proposed as follows:

$$\overline{\lambda}_{i} = \frac{|\lambda_{i}|}{\sqrt{\lambda_{f}^{2} + \lambda_{o}^{2} + \lambda_{p}^{2} + \lambda_{d}^{2}}} \quad i \in \{f, o, p, d\}$$
(2)

The experimental results show that the ratio of the four parameters of the receptive field is approximately  $\bar{\lambda}_f: \bar{\lambda}_o: \bar{\lambda}_p: \bar{\lambda}_d = 4:4:1:1$ . The experimental results approximately show the contribution degrees and sizes of the four parameters.

### Improvement 2: Feature extraction method based on MyTICA algorithm

The myTICA [10] method was used to replace the PCA method in the original SFA algorithm to extract the slow features of natural images, which effectively solved the problem of the original SFA algorithm that the features did not have independence and do not contain the visual spatial topological relationships. Inspired by the theory of visual selectivity, the basis functions in the family of basis functions *Visual\_sub\_set(i)* have the global and local visual selectivity obeying the rule of "similar when being close", and form a stable spatial topological structure with close distribution. Both the slow information and the slow topology exhibit the visual selection consistency continuity. *Visual\_sub\_set* is simulated and expressed with Gabor functions as follows:

$$Visual\_sub\_set = {Visual\_sub\_set(p\_add,i) / p\_add = 1,2,3,...,N}, (3) i \in \{\lambda_f, \lambda_o, \lambda_p, \lambda_d\}$$

where, *f* represents the frequency of the receptive field, o the orientation of the receptive field, p the phase of the receptive field, d the visual spatial position of the receptive field, and N the number of elements in the basis function family.

#### Improvement 3: Feature basis expansion algorithm based on visual selection consistency continuity (VSCC FBEA)

The basis space expansion of the original SFA algorithm adopted the polynomial expansion method, which has high computational complexity, lacks the support of computational theory of vision for primates, and is also unable to learn the short-term and long-term features with visual selection consistency continuity. What is more, the generated basis function space is not over-complete. Inspired by the visual selection consistency continuity calculation theory, this paper used the Gabor basis functions to model the visual receptive field. In order to obtain the long- and short-term features with visual selection consistency continuity from the frequency, orientation, phase and position of each element in the Gabor basis function family, the deep LSTM algorithm was used to replace the original polynomial expansion method, which realized the nonlinear expansion of the basis functions in the visual basis function space and their topological structure, and helped generate an over-complete basis function family and its spatial topological structure with visual selection consistency continuity.

Inspired by the theory of visual selection consistency continuity, this paper proposed the LOPM algorithm to optimize the invariant feature forest, and obtained the optimized over-complete basis function family based on visual selection consistency continuity. The detailed implementation is as follows:

#### Improvement 3.1 Structure of the VSCC\_FBEA algorithm

Based on the above innovations, the flow chart of the feature basis expansion algorithm used in this paper is as follows Figure 1:



Figure 1. Flow chart of the VSCC\_FBEA algorithm

where, the basis function space is formed by the basis functions of the natural image, which represents the visual spatial selectivity with respect to natural images, and is expressed Visual\_sub\_set = as { $Visual\_sub\_set(p\_add, i)|p\_add = 1, 2, 3, ..., N$ },  $i \in$  $\{\lambda_f, \lambda_o, \lambda_p, \lambda_d\}$ ; inspired by the visual selectivity rule of primates "similar when being close and different when being distant", the basis function space is divided into basis function subspaces, expressed as  $Visual\_sub\_set(p\_add) =$ {*Visual\_sub\_set*( $p_add, \lambda_i$ )| $p_add = 1, 2, 3, ..., N$ ;  $i \in$  $\{\lambda_f, \lambda_o, \lambda_p, \lambda_d\}\};$  based on the different contributions of the parameters in the basis functions, the basis function subspace Visual sub set(p add) is divided into four subsets -Visual sub set(p\_add,  $\lambda_f$ , Visual sub set(p add,  $\lambda_o$ ), *Visual sub set*(p add,  $\lambda_p$ ) and *Visual sub set*(p add,  $\lambda_d$ ). The counter *p* add represents the dynamic curser of the basis function space, and num1 the number of elements in the basis function space. L sample means sampling from the set Visual sub set(p add, i) using the random sampling method according to the sample size L, where L is less than or equal to the number of elements in the set.

#### VSCC\_FBEA algorithm

Feature basis expansion algorithm based on visual
selection consistency continuity (VSCC_FBEA)
Input:
$Visual\_sub\_set = {Visual\_sub\_set(p\_add, i) p\_add =$
$1,2,3,\ldots,N\}, i \in \{\lambda_f,\lambda_o,\lambda_p,\lambda_d\}$
Output: $(\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d)$

#### Begin:

1. By the L\_sample sampling method, take a subset of *Visual\_sub\_set*(*p\_add*, *i*) from the *Visual\_sub\_set* normalized.

Divide Visual sub set(p add, 2. i) into Visual sub set(p add,  $\lambda_f$ ), Visual sub set(p add,  $\lambda_o$ ), *Visual sub set*(p add,  $\lambda_p$ ) and *Visual sub set*(p add,  $\lambda_d$ ). 3.  $(\lambda_{f_{1}}, \lambda_{o}, \lambda_{b}, \lambda_{d}) = (F \text{ LSTM } (Visual sub set(p add, \lambda_{f}))),$ O LSTM (Visual sub set(p add,  $\lambda_o$ )), P LSTM (Visual sub set(p add, D LSTM  $\lambda_p$ )), (Visual sub set( $p_add$ ,  $\lambda_d$ ))).

4. By the formula (2), Realize parameter standardization based on visual selectivity,  $(\lambda_f, \lambda_o, \lambda_p, \lambda_d) \rightarrow (\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d)$ . End

#### Improvement 3.2 Basis Fusion Algorithm (BFA algorithm)

There is a certain dependence between the four parameters of the *Gabor* basis function generated through prediction. Therefore, the autocorrelation and cross-correlation coefficient matrix is calculated as follows:

$$\delta = \begin{bmatrix} w_{ff} & w_{fo} & w_{fp} & w_{fd} \\ w_{of} & w_{oo} & w_{op} & w_{od} \\ w_{pf} & w_{po} & w_{pp} & w_{pd} \\ w_{df} & w_{do} & w_{dp} & w_{dd} \end{bmatrix}$$
(4)

This matrix describing the correlation between the four visual parameters can be used to realize the fusion of *Gabor* predictive parameters. The fusion algorithm (BFA algorithm) is as follows:

$$(\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d) = (\lambda_f, \lambda_o, \lambda_p, \lambda_d) \times \delta$$
(5)

The four parameters  $(\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d)$  generated by the BFA algorithm have learned not only the visual consistency of the basis function family, but also the correlation between the parameters. In this way, the consistency and correlation of the basis functions are integrated.

Basis Fusion Algorithm (BFA):
Basis Fusion Algorithm (BFA)
Input: $(\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d), \delta$
Output: $(\hat{\lambda}_f, \hat{\lambda}_o, \hat{\lambda}_p, \hat{\lambda}_d)$
Begin:
$(\bar{\lambda}_f, \ddot{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d) = (\lambda_f, \lambda_o, \lambda_p, \lambda_d) \times \delta$
End

### Improvement 3.3 *Lipschitz-Orthogonal-Pruning-Method* (LOPM algorithm)

Inspired by visual selectivity, among the four parameters of the Gabor function, frequency and orientation have greater contributions and play the absolute role in the Gabor function. Perform Lipschitz [25-30] conditional calculations on the frequency and orientation parameter sequences.

The frequency sequence is *frequency\_set* =  $\{\lambda_f^1, \lambda_f^2, \lambda_f^3, \dots, \lambda_f^M, \dots, \lambda_f^N\}$ ; and the orientation sequence is *oritation\_set* =  $\{\lambda_o^1, \lambda_o^2, \lambda_o^3, \dots, \lambda_o^M, \dots, o_o^N\}$ , where M represents the number of basis functions obtained through myICA, and from M+1 to N are the expanded basis function generated by the algorithm proposed in this paper; the parameter generated by the LSTM algorithm  $\alpha = \{\lambda_f^*, \lambda_o^*, \lambda_p^*, \lambda_d^*\}$ , and the calculated mean and variance  $(\mu_f, \delta_f)$  of the samples from *frequency set* are respectively the

Lipschitz coefficient L of this sequence and the constraint value of the coefficient. The calculated mean and variance ( $\mu_o$ ,  $\delta_o$ ) of the samples from *Orientation\_set* are respectively the Lipschitz coefficient L of this sequence and the constraint value of the coefficient.

The LOPM algorithm is described as follows:

When  $\alpha$  satisfies  $min(-(\mu_f - \delta_f), -(\mu_o - \delta_o)) \le \lambda_f^* + \lambda_o^* \le min((\mu_f - \delta_f), (\mu_o - \delta_o))$ , and  $\forall \beta \in Gabor\_set$ , then, if  $\langle \alpha, \beta \rangle \le \varepsilon$ , the vectors  $\alpha$  and  $\beta$  are similar, and perform the pruning operation, and keep  $\beta$ . Otherwise, add  $\alpha$  to the set  $Gabor\_set$ . The pruning coefficient  $\varepsilon$  is an arbitrary small positive number that controls the complexity of pruning. Lipschitz-Orthogonal-Pruning- Algorithm (LOPM):

 $\delta_o$ ))  $\leq \lambda_f^* + \lambda_o^* \leq min((\mu_f - \delta_f), (\mu_o - \delta_o))$ , if  $\langle \alpha, \beta \rangle \leq \varepsilon$ , then perform pruning, or else push  $\alpha$  into the basis function subspace. End

### **Improvement 4:** *Feature expression method based on visual invariance theory (visual invariance theory -FEM)*

The original SFA algorithm only extracts the PCA features of a natural image, which reflect the contribution of each component of the image in the whole image, but cannot show the visual consistency of the information in the visual space or the invariance of the two spatial structures of the visual information. The two spatial structures include the spatial topological invariance of each basis function family in the visual space and the topological invariance of the spatial distribution of the basis functions in the visual space.

Based on this, this paper proposed the visual invariance theory –FEM method. This is a feature expression method that can describe the visual information and the topological invariance of the information space, established based on the basis function families generated in step 1, 2, and 3 and according to the visual selection consistency continuity theory. The specific implementation process is as follows:

Basis function families with visual selection consistency continuity can be generated from each natural image -Visual sub set={Visual sub set{p add, p\_add *i*}| =1,2,3,...N;  $i \in \{\lambda_f, \lambda_o, \lambda_p, \lambda_d\}$ , where Visual sub set(p add, i) contains N function families. According to the four parameters of visual selectivity, each function family has distinguishable visual features and visual spatial topological structures, which reflect the overall features of the global visual space; and the basis functions within each function family in the visual space also contain the visual information of the basis functions and the information of the spatial topological structure between the functions. The above-mentioned global and local visual information and the spatial topological structure information all have visual selection consistency continuity, thus forming the features with visual selection consistency continuity.

Therefore, based on the theory of visual selection

consistency continuity and the theory on primates' visual information processing – "parallel operation on the global scale and serial operation on the local scale", a feature forest expressing the slow features and slow topologies extracted by the algorithm proposed in this paper is constructed as the following Figure 2:



Figure 2. Natural image feature forest with visual selection consistency continuity

where, *Visual\_sub\_set*(*p\_add*, *i*) corresponds to the p\_add-th basis function family in the visual space of the natural image, representing the local visual information of the basis function family *Visual\_sub\_set*(*p\_add*, *i*). For *Visual\_sub\_set*(*p\_add*, *i*), the extraction of features between basis function families is a parallel process, while the extraction of features within the same family is a serial process, which is conducive to the distributed parallel design of the algorithm. The function family *Visual\_sub\_set* expresses the global features of a natural image, including the global visual information and the global topological structure information.

In summary, the slow features of a natural image are expressed as:

$$\begin{cases} Visual\_sub\_set\_expression=T\_encode(Visual\_sub\_set) \\ =(index(p\_add, j), gabor(p\_add, j)) \\ index(p\_add, j)=T1\_encode(Visual\_sub\_set(p\_add, i,k)) \\ gabor(p\_add, i)=T2\_encode(Visual\_sub\_set(p\_add, i,k)) \end{cases}$$
(6)

where,  $p\_add$  indicates the visual space basis; *j* the position of the element in the visual space basis;  $T\_encode$ -{ $T1\_encode$ ,  $T2\_encode$ }; and the parameter k the basis function cursor in the basis function family. The specific coding rules are as follows:

Feature	expression	method	based	on	visual	invariance
theory (visi	ual invarian	ice theory	-FEM	):		
Fostuno	annuagaian	mathad	bagad			manianaa

redure expression method bused on visual invariance
theory (visual invariance theory –FEM)
Input: Visual sub set(p add, i)
Output: index, GM_gabor
Begin:
//T1_encode:
1.Get $\lambda_0$ from <i>Visual_sub_set</i> ( <i>p_add</i> , <i>i</i> ),
$2.\mathrm{if}(\lambda_0)$
$index(p\_add, int(\lambda_o))=1;$
else
$index(p_add, int(\lambda_o))=0$
end
//T2 encode:

If $(index(p_add, int(\lambda_o))=1)$
GM_gabor(p-add,j)= Visual_sub_set(p_addi,k);
GM_gabor(p-add,j)
else
GM_gabor(p-add,j)=0;
end
end

After the T\_encode transformation, the generated matrixes *Index* and *GM\_gabor* are expressed as following the Table 2:

 Table 2. Feature matrix Index (with 360 features per row)

Index	Value								
Index(1,j)	1	0	0		1				
Index(2,j)	0	0	1		0				
Index(3,j)	1	1	0		1				
•									
Index(n,j)	0	1	1		1				

The matrix *Index* has frequency and orientation selectivity in the columns and rows, respectively, expressed in an ascending order, and the range of the orientation is specified as  $[1^\circ, 360^\circ]$ . Therefore, the features are described as the Table 3:

 Table 3. Receptive field corresponding to the feature matrix

 (GM\_gabor)

chain			value	
chain(1)	g11	g12	g13	 g1360
chain(2)	g11	g12	g13	 g1360
chain(3)	g11	g12	g13	 g1360
•••				
chain(n)	g11	g12	g13	 g1360

In the feature matrix *Index*, "1" and "0" indicate whether there is a receptive field at the location; the value "1" indicates that there is a receptive field, and there is a corresponding function Gabor(i,j) at the node of Index(i,j). These functions form the feature matrix  $(GM\_gabor)$ , as shown in Table 3.

#### 4. ALGORITHM DESIGN

#### 4.1 Structure of the algorithm

With the above innovations, the flow chart of the proposed basis function expansion algorithm is shown as the following Figure 3:



Figure 3. Flow chart of the whole algorithm

where, *image set* is the natural image database. The basis function family presenting the visual spatial selectivity of the natural image is denoted as Visual\_sub\_set = { $Visual\_sub\_set(p\_add, i)$ | $p\_add = 1, 2, 3, ..., N$ },  $i \in$  $\{\lambda_f, \lambda_o, \lambda_p, \lambda_d\}$ , as generated by the *MyTICA* algorithm; inspired by the visual selectivity rule of primates "similar when being close and different when being distant", the basis function space is divided into basis function subspaces of the natural image, expressed as Visual sub set =  $\{Visual\_sub\_set(p\_add,i) | p\_add = 1,2,3,\ldots,N\}, i \in$  $\{\lambda_f, \lambda_o, \lambda_p, \lambda_d\}$ ; based on the different contributions of the parameters in the basis functions, the basis function subspace Visual sub set(p add, i) is divided into four subsets -Visual sub set(p add,  $\lambda_f$ , Visual sub set(p add,  $\lambda_o$ ), Visual sub set(p add,  $\lambda_p$ ) and Visual sub set(p add,  $\lambda_d$ ). The counter p add represents the dynamic curser of the basis function space, and num1 the number of elements in the basis function space. L sample means sampling from the set Visual\_sub\_set(p\_add, i) using the random sampling method according to the sample size L, where L is less than or equal to the number of elements in the set.

#### 4.2 Steps of the feature extraction algorithm

1. Based on the natural image database *image\_set*, construct the primate visual space basis function family  $Visual\_sub\_set = \{Visual\_sub\_set(p\_add, i) | p\_add =$  $1,2,3,...,N\}, i \in \{\lambda_f, \lambda_o, \lambda_p, \lambda_d\}$  using the myICA method, initialize number1=N, and control variable p\_add=1, and initialize the correlation coefficient matrix  $\delta$ .

2. Normalize each subset in the basis function family *Visual\_sub\_set*.

3. Use the L\_sample sampling method to take out the four subsets *Visual\_sub\_set*{*p\_add*,  $\lambda_{ff}$ }, *Visual\_sub\_set*{*p\_add*,  $\lambda_{of}$ }, *Visual\_sub\_set*{*p\_add*,  $\lambda_{of}$ }, *Visual\_sub\_set*{*p\_add*,  $\lambda_{af}$ }, from the normalized *Visual\_sub\_set*, input the four subsets sequentially into the F\_LSTM, O\_LSTM, P\_LSTM and D\_LSTM algorithms to obtain the visual selectivity parameters  $\lambda_{fs}$ ,  $\lambda_{o}$ ,  $\lambda_{p}$ ,  $\lambda_{d}$ . In this way, the prediction of visual selectivity parameters is realized.

4. Normalize the parameters with visual selectivity according to the parameter contribution algorithm rule (2) and obtain the normalized parameters  $\lambda_f$ ,  $\lambda_o$ ,  $\lambda_p$ ,  $\lambda_d$ .

5. Substitute  $(\lambda_f, \lambda_o, \lambda_p, \lambda_d)$  into formula (5) according to the BFA algorithm to calculate the corresponding fusion parameters  $(\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d)$ , and realize the fitting of the Gabor function.

6. Output the parameters of the fitted Gabor basis functions to  $Visual\_sub\_set\{p\_add, i\}$ , execute the LOPM algorithm, and insert the unpruned Gabor basis functions into the relevant database to expand the basis functions of the slow feature algorithm.

7. p\_add++;

8. If  $p\_add \le number1$ , the algorithm goes to step 3, and continues the execution. Otherwise, the algorithm stops, and the generation process of slow feature prediction is completed.

9. Use the *visualinvariancetheory-FGM* algorithm to generate the feature of the invariant data and the corresponding feature forest *Visual sub set expression*.

*Feature expression algorithm based on visual invariance theory* (BFA):

Feature expression algorithm based on visual invariance theory (BFA)

Input: image\_set Output: Visual sub set expression

Begin:

1. Based on the natural image database image\_set, construct the primate visual space basis function family Visual\_sub\_set = {Visual\_sub\_set(p\_add,i)|p\_add = 1,2,3,...,N},  $i \in {\lambda_f, \lambda_o, \lambda_p, \lambda_d}$  using the myICA method, initialize number1=N, and control variable p\_add=1, and initialize the correlation coefficient matrix  $\delta$ .

2. Perform normalization on each subset of the basis function family Visual\_sub\_set.

3. Use the L\_sample sampling method to take out four subsets Visual\_sub\_set{p\_add,  $\lambda_{ff}$ , Visual\_sub\_set{p\_add,  $\lambda_{of}$ , Visual\_sub\_set{p\_add,  $\lambda_{p}$ } and Visual\_sub\_set{p\_add,  $\lambda_{d}$ } from the normalized Visual\_sub\_set; input sequentially the four subsets into the F\_LSTM, O\_LSTM, P\_LSTM and D\_LSTM algorithms to obtain the visual selectivity parameters  $\lambda_f$ ,  $\lambda_o$ ,  $\lambda_p$ ,  $\lambda_d$  to realize the prediction of visual selectivity parameters.

4. Normalize the parameters with visual selectivity according to the parameter contribution algorithm rule (2) and obtain the normalized parameters  $\lambda_f$ ,  $\lambda_o$ ,  $\lambda_p$ ,  $\lambda_d$ .

5. Substitute  $(\lambda_f, \lambda_o, \lambda_p, \lambda_d)$  into formula (5) according to the BFA algorithm to calculate the corresponding fusion parameters  $(\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d)$  and realize the fitting of the Gabor function.

6. Output the parameters of the fitted Gabor basis functions to Visual\_sub\_set{p\_add, i}, execute the LOPM algorithm, and insert the unpruned Gabor basis functions into the relevant database to expand the basis functions of the slow feature algorithm.

7. *p* add++;

8. If  $p_{add} \leq number1$ , the algorithm goes to step 3, and continues the execution. Otherwise, the algorithm stops, and the generation process of slow feature prediction is completed.

9. Use the visualinvariance theory-FGM algorithm to generate the feature of the invariant data and the corresponding feature forest Visual\_sub\_set\_expression. End

#### 4.3 Evaluation of the algorithm

In order to evaluate the prediction ability and universality of the algorithm proposed in this paper, three evaluation methods, namely root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE), were used. The evaluation methods are presented as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{N} (y_i - y_i^*)^2}$$
(7)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_i^*|$$
(8)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - y_i^*|^2}{y_i^*}$$
(9)

where,  $y_i^*$  is the basis function generated by the myICA algorithm;  $y_i$  is the basis function predicted by the proposed algorithm, and n is the number of samples for observation.

This paper adopted the classification method proposed in literature [11].

#### 5. TESTS AND CONCLUSION ANALYSIS

#### 5.1 Generation and display of the test gallery

According to the above theoretical analysis, the datasets for study of the parameter setting and performance of the proposed algorithm were selected from the dataset LSVRC2012. 300 classes are randomly selected from the dataset LSVRC2012, and 899 pictures were randomly selected for each class, as the training set LSVRC201\_train; and the remaining 401 ones were used as the test set. Figure 4 shows some sample images selected from the gallery.



Figure 4. Sample pictures from the INRIA Holidays gallery for the test

#### 5.2 Consistency of the predicted basis function

The sampling requirements are as follows: (1) the sampling size is  $11 \times 11$  pixels; (2) the frequency is 2048 times; and (3) the method is random sampling. According to these requirements, a subset was selected from LSWRC201\_train, which contained 899 pictures, and samples were taken from each image and formed the image set *original\_set*. Through the execution of the algorithm [10], the visual space basis function family *Visual\_sub\_set* was generated. And through the execution of the proposed VSCC\_FBEA algorithm and LOPM algorithm, the basis function family was optimized into *Visual\_sub\_set* {*p\_add*, *i*}. The consistency features of the four visual selectivity parameters of the Gabor function were tested, with the results shown the Figure 5:

According to the results, the optimized basis function family  $Visual\_sub\_set{p\_add, i}$  was generated based on the natural image set *original\\_set* by the proposed algorithm; and the consistency distribution of the four visual parameters of the basis function Gabor(f,o,p,d) with visual selectivity in the function family is shown in a) and b) of Figure 5. It can be seen that the basis function predicted and the initial basis function generated by the algorithm [10] have consistency continuity and are closely distributed in space. The basis functions in different basis function families are obviously distinguishable, with no visual selection consistency continuity. This is because, for the natural image set, the four visual parameters of the basis function family generated by the proposed

algorithm have consistency continuity, which is manifested as spatial agglomeration in the image. Therefore, the test results are consistent with the theory.



a) Consistency distribution of frequency, orientation and



b) Consistency distribution of frequency, orientation and position

Figure 5. Visual consistency distribution of the predicted basis function

### 5.3 Effect of L in L\_sample on the prediction performance of the algorithm

In the proposed algorithm, the L\_sample method consists of two parts - the value of L and the sampling method. Specifically, the value of L directly determines the number of basis functions to be sampled, and determines the nature and number of basis functions to be generated; the sampling method is related to the quality of the samples. Therefore, L\_sample has significant effects on the prediction and recognition performance of the algorithm. To evaluate the prediction of the proposed algorithm with respect to the dataset LSVRC201\_train, different L values were used, and the evaluation results by different methods are shown the following Table 4.

According to the test results, after the sampling method was determined, when the value of L exceeded half of the size of the basis function family *Visual\_sub\_set{p\_add, i}*, that is, when L>size(*Visual\_sub\_set{p\_add, i}*)/2, as the value of L increased, the three evaluation indicators RMSE, MAE and MAPE decreased. The primary reason is that the increase of the L value led to an increase in the number of basis functions with visual selection consistency continuity, enhancing the learned visual selectivity features and resulting in stronger visual selection consistency continuity of the basis functions generated. In this way, the errors between basis functions decreased. Therefore, the three error-based evaluation indicators decrease with the increase of the L value.

Table 4. Relationship between the L value and the prediction performance of the algorithm

Evaluation method	L	N/2	$\frac{7}{16}N$	$\frac{5}{8}N$	$\frac{11}{16}N$	$\frac{3}{4}N$	$\frac{13}{16}N$	$\frac{7}{8}N$	$\frac{15}{16}N$	Ν
RMSE		15.13	14.38	13.52	13.15	12.19	11.74	9.18	9.02	8.36
MAE		7.81	7.23	7.05	6.45	6.18	6.03	5.79	5.56	5.31
MAPE		9.72	9.25	9.03	8.76	8.29	7.46	7.02	6.64	6.01

After the sampling method was determined, the change of the L value led to the change in the constraint on the generation of the basis functions, and thus the visual selection consistency continuity of the basis functions inevitably changed accordingly. When the test conditions were determined as follows:

The value range of L is [N/2,N], the step size is (NN/2)/11, and A(N)=N/2+(n-1)\*(NN/2)/11; when N =2888, with the proposed algorithm executed on the dataset LSVRC201\_train, the performance results of the algorithm are given as the following Figure 6 based on different L values:



Figure 6. Effect of the L value on the recognition rate of the algorithm

According to the test results, after the L-sample method was determined, with the increase of the L value, the recognition rate of the algorithm first increased and then gradually stabilized. When the L value was 2346, which means 2346 basis functions were involved in the prediction of visual selection consistency continuity (accounting for 81% of the total), the recognition performance of the algorithm approached 98.99% and tended to be stable. This indicates that, when basis functions involved in the prediction accounted for 80% or more of the elements in the basis function family, the proposed algorithm will achieve the best prediction performance. The reason is that when the proportion of basis functions exceeds 80%, the basis functions generated by the proposed algorithm through learning have very strong visual selection consistency, which increases the recognition rate of the algorithm.

In summary, when the proposed algorithm uses random sampling and the initial value of the sampling coefficient L is [0.5,1], the algorithm will have good recognition performance. In particular, when the value of L is greater than 80% of the number of elements in the basis function family, the algorithm will have a recognition rate of more than 98.99%. The above parameters can be used as the constraints for subsequent tests.

# 5.4 Effect of the pruning control parameter ( $\varepsilon$ ) on the spatial distribution of basis functions and the recognition performance of the algorithm

The pruning control parameter  $\varepsilon$  determines whether the generated basis functions can be added to the generated basis

function family, and accordingly, determines the overcompleteness of the function family. Therefore, the parameter  $\varepsilon$  is essential to the prediction and recognition performance of the proposed algorithm. With different values of the parameter  $\varepsilon$ , the prediction and recognition performance of the proposed algorithm with respect to the dataset LSVRC201\_train were tested, as shown the following Figure 7:



a) Relationship between the pruning coefficient  $\varepsilon$  and the number of elements in the generated basis function family



b) Relationship between the pruning coefficient  $\varepsilon$  and the recognition performance of the algorithm

Figure 7. Relationship between the pruning coefficient  $\varepsilon$  and the performance of the proposed algorithm

As can be seen, as the pruning control parameter  $\varepsilon$  increased, the number of elements in the generated basis function family increased, and the recognition performance of the algorithm also increased, but when  $\varepsilon \ge 0.12$ , the number of elements in the base function family generated by the algorithm tended to be flat, and the recognition performance of the algorithm gradually stabilized at above 97%. The primary reason is that, as the pruning control parameter  $\varepsilon$  increased, the orthogonal condition gradually relaxed, and the size of the basis function family increased accordingly. The more basis functions involved in prediction, the greater the over-completeness of *Visual\_sub\_set{p\_add, i}*; and the more basis functions participating in the recognition training of the algorithm, the stronger the visual selection consistency continuity, and the better the recognition performance of the algorithm.

In summary, when  $\varepsilon \ge 0.12$ , the proposed algorithm can achieve the best prediction and recognition performance, with a recognition rate of 97.83%.

#### 5.5 Feasibility evaluation of the proposed algorithm

In summary, with the respect to the training set LSVRC201\_train, if the random sampling method is used, and the sampling parameter L is greater than 80% of the number of elements in the basis function family, and when the pruning control parameter  $\varepsilon \ge 0.12$ , the proposed algorithm can achieve the best recognition performance. The intra-class and interclass Bhattacharyya distances are calculated, as shown in the following Figure 8.



Figure 8. Intra-class and inter-class distributed distances

According to the ROC (receiver operating characteristic) curve, when the test threshold is 8, the algorithm achieved a recognition rate of 99.66%, and both the false recognition rate and the false rejection rate are less than 0.33%. This shows that the proposed algorithm is feasible and has strong classification capacity.

### 5.6 Comparison of the recognition performance of different algorithms

Based on the training set LSVRC201\_train, by the random sampling method and with the sampling parameter L being

greater than 80% of the number of elements in the basis function family and the pruning control parameter  $\varepsilon \ge 0.12$ , the performance of the proposed algorithm was compared with that of the other algorithms, including SFA [1], GSFA [8], TICA [9], myTICA [10] and mySFA [11]. The results are shown in Figure 9:



Figure 9. Comparison of the recognition performance of different algorithms

According to the test results, with respect to the training set LSVRC201 train, with the increase of sampling window size and sampling frequency, the size of the sample set also increased and the recognition rates of all algorithms first increased and then tended to be flat. When the sampling parameter was in the range of [7, 13], the performance of various algorithms tended to be stable, and the proposed algorithm showed a higher recognition rate than the others. The reason is that the basis functions generated by the proposed algorithm have visual selection consistency. The basis functions in the same basis function family have visual selection consistency continuity while those in different families do not, which is in line with the visual selection feature of "adjacent within one class and distant between classes". In addition, the proposed algorithm extracts the basis functions, basis function families and their spatial topological structures in natural images, and constructs invariant information and invariant topological features in the visual space. So in essence, the above-mentioned invariant visual features are superior to those used in other algorithms. When the sampling window is too small, the visual feature information and structures presented by the basis functions extracted by the proposed algorithm will be incomplete, and thus cannot reflect the visual invariance of the basis functions; when the window is too large, the extracted area will cover multiple receptive fields, and thus the extracted basis functions will fuse the visual features of multiple receptive fields, leading to chaotic feature information of the basis functions: and when the sampling frequency is low, the visual features of natural images cannot be fully presented, and thus the information of the receptive fields is lost. All the above reasons will ultimately reduce the recognition performance of the algorithm.

In this test, the sample set was the result of optimization. The mapping relationship for the parameter is shown as the following Table 5:

1	2	3	4	5	6	7	8	9	10	11	12
(3*3,25	(3*3,51	(3*3,10	(5*5,10	(5*5,20	(5*5,40	(7*7,10	(7*7,20	(7*7,40	(11*11,20	(11*11,40	(13*13,40
6)	2)	24)	24)	48)	96)	24)	48)	96)	48)	96)	96)

**Table 6.** Anti-noise capacity of the algorithm

Intensity type	0	1	3	5	7	9	11	13	15
Speckl	99.91	99.21	98.71	97.34	91.11	87.21	81.61	71.12	40.21
Salt & Pepper	99.88	99.11	98.10	95.42	92.05	87.41	76.29	71.72	45.56
Gaussian	99.99	98.81	98.01	94.36	92.08	86.92	77.42	66.55	42.43

#### 5.7 Analysis on the anti-noise capacity of the algorithm ( $\sqrt{}$ )

The anti-noise performance is an important indicator of the robustness of an algorithm. Generally speaking, an algorithm should be able to resist different types or intensities of noise, which is also an important prerequisite for the application of the algorithm. In this paper, the type and intensity of noise have significant effects on the prediction and recognition performance of the proposed algorithm, as different noise types (Gaussian, Salt&pepper, and Speckl) and intensities have different effects on the learning of the visual selection consistency continuity of the generated basis functions. If there is too much noise, the generated basis functions will not have visual selection consistency continuity, which is not worth studying. Therefore, the noise intensity was set at [0,1,3,5,7,9,11,13,15] in the test, with % as the unit. Based on the training set LSVRC201 train, by the random sampling method and with the sampling parameter L being greater than 80% of the number of elements in the basis function family and the pruning control parameter  $\varepsilon \ge 0.12$ , the results of antinoise capacity test was displayed as the Table 6.

According to the test results, when the noise intensity was in the range of [0, 15%], the recognition rate of the algorithm decreased as the noise intensity increased. When the noise intensity was between [0, 5%], the recognition rate of the algorithm decreased slowly as the noise intensity increased; and when it exceeded this range, the recognition rate decreased rapidly as the noise intensity increased. In terms of noise types, the functional expression of Gaussian noise is a subset of the Gabor function, which has obvious visual "ON-OFF" selectivity. Therefore, the increase of Gaussian noise will have two effects: (1) increasing the basis functions; (2) changing the features of the basis functions. It will directly result in the inability of the extracted basis functions to reflect the essential features of natural images. However, the visual selection consistency pruning technology of the proposed algorithm is able to eliminate these basis functions, so the increase of this type of noise has little effect on the overall performance of the algorithm. Salt & pepper noise, featured by black and white dots, has minor visual "ON-OFF" selectivity. This type of noise can change the visual features of a natural image within a region, thereby affecting the regional learning ability of the Gabor basis function. With the increase of the sampling window size, the impact will gradually become smaller, but the pruning technology used in the proposed algorithm has no obvious mitigating effect on this type of noise. Unlike the previous two types of noise, Speckl noise has larger monomer structures and can independently form Gabor basis functions with visual selectivity, so it can change the visual features and visual spatial topological structures of a natural image within a region in a fused manner. The visual features extracted by the proposed algorithm are fused features instead of the visual features of natural images. Therefore, this type of noise has great impact on the recognition rate of the proposed algorithm. In summary, under the same noise intensity, the sequence of the noise types is Speckl, Salt&pepper and Gaussian noise in terms of their impacts on the recognition rate of the proposed algorithm from high to low.

#### 6. CONCLUSIONS

The traditional slow feature analysis (SFA) algorithm has no support of computational theory of vision for primates, nor does it have the ability to learn the global features with visual selection consistency continuity. And what is more, the algorithm is highly complex. Based on this, Slow Feature Extraction Algorithm Based on Visual selection consistency continuity and Its Application was proposed. This paper solved the drawbacks of the traditional SFA algorithm through four innovations: 1. it replaced the PCA method of the traditional SFA algorithm with the myTICA method, extracted the Gabor basis functions showing the invariance of natural images, and generated the basis function family; it used the VSCC FBEA algorithm to realize the basis function expansion of the features with visual selection consistency continuity; 3. it proposed using the LOPM algorithm to optimize the overcomplete basis function family, and generated the optimized over-complete basis function family; 4. It adopted the visual invariance theory -FEM to generate the set of invariable features containing the visual information of natural images and its spatial topological relations. Through experiments, this paper established the method to obtain and optimize the algorithm parameters and also obtained the optimized parameters for the LSVRC2012 data set; based on this, the proposed algorithm was compared with the SFA, GSFA, TICA, myICA and mySFA algorithms, which shows that the proposed one is correct and feasible; when the classification threshold of the algorithm was set at 8.0, the recognition rate of the proposed algorithm reached 99.66%, and neither of the false recognition rate and the false rejection rate was higher than 0.33%. The proposed algorithm has good performance in prediction and classification, and also exhibits good anti-noise capacity under limited noise conditions.

At this stage, the research only explored the application of the proposed algorithm in the analysis and understanding of natural images; in the future, further studies will be carried out to explore the application of the proposed algorithm in signal recognition, motion capture, weather forecast and image sequence analysis.

#### ACKNOWLEDGMENT

This work is supported by the Key R&D Projects in Hebei

Province of China (Grant No.: 20310301D), the Key R&D Projects in Hebei Province of China (Grant No.: 19210111D), the social science foundation of Hebei province of China (Grant No.: HB18TJ004), Science and technology planning project of Hebei Province of China (Grant No.: 15210135), National Science and Technology Infrastructure Program (Grant No.: 2015BAH43F00). the Key Research and Development Program of Hebei province of China(20370801D).

#### REFERENCES

- [1] Hinton, G.E. (1990). Connectionist learning procedures. Machine Learning, 3: 555-610. https://doi.org/10.1016/B978-0-08-051055-2.50029-8
- Wiskott, L., Sejnowski, T.J. (2002). Slow feature analysis: Unsupervised learning of invariances. Neural Computation, 14(4): 715-770. https://doi.org/10.1162/089976602317318938
- [3] Berkes, P., Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. Journal of Vision, 5(6): 9. https://doi.org/10.1167/5.6.9
- [4] Franzius, M., Sprekeler, H., Wiskott, L. (2007). Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS Computational Biology, 3(8): e166. https://doi.org/10.1371/journal.pcbi.0030166
- [5] Franzius, M., Wilbert, N., Wiskott, L. (2008). Invariant object recognition with slow feature analysis. In International Conference on Artificial Neural Networks, 5163: 961-970. https://doi.org/10.1007/978-3-540-87536-9\_98
- [6] Klampfl, S., Maass, W. (2009). Replacing supervised classification learning by Slow Feature Analysis in spiking neural networks. Advances in Neural Information Processing Systems, 22: 988-996.
- [7] Ma, K.J., Han, Y.J., Tao, Q., Wang, J. (2011). Kernelbased slow feature analysis. Pattern Recognition and Artificial Intelligence, 24(2): 153-159. https://doi.org/10.3969/j.issn.1003-6059.2011.02.001
- [8] Luciw, M., Kompella, V.R., Schmidhuber, J. (2012). Hierarchical incremental slow feature analysis. In Workshop on Deep Hierarchies in Vision.
- [9] Escalante-B, A.N., Wiskott, L. (2012). Slow feature analysis: Perspectives for technical applications of a versatile learning algorithm. KI-Künstliche Intelligenz, 26(4): 341-348. https://doi.org/10.1007/s13218-012-0190-7
- [10] Zhao, Y.M., Fang, J., Ji, S.J. (2015). Feature extraction algorithm based on complex visual information of natural image and its application. Computer Applications and Software, 32(11): 200-205. https://doi.org/10.3969/j.issn.1000-386x.2015.11.047
- [11] Zhao, Y.M. (2019). Design and application of an adaptive slow feature extraction algorithm for natural images based on visual invariance. Traitement du Signal, 36(3): 209-216. https://doi.org/10.18280/ts.360302
- [12] Zhang, J.H., Zhu, Q., Song, L. (2019). A wavelet-based self-adaptive hierarchical thresholding algorithm and its application in image denoising. Traitement du Signal, 36(6): 539-547. https://doi.org/10.18280/ts.360609
- Zhao, Y.M. (2020). Slow variation feature extraction algorithm for defocused image sequences based on visual selectivity. Computer Applications and Software, 37(1): 205-212. https://doi.org/10.3969/j.issn.1000-

386x.2020.01.034

- [14] Zhao, Y.M. (2020). Improvement and application of multi-layer LSTM algorithm based on spatial-temporal correlation. Ingénierie des Systèmes d'Information, 25(1): 49-58. https://doi.org/10.18280/isi.250107
- [15] Zhao, Y.M. (2020). Spatial-temporal correlation-based LSTM algorithm and its application in PM2.5 prediction. Revue d'Intelligence Artificielle, 34(1): 29-38. https://doi.org/10.18280/ria.340104
- [16] Koch, P., Konen, W., Hein, K. (2010). Gesture recognition on few training data using slow feature analysis and parametric bootstrap. In the 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. https://doi.org/10.1109/IJCNN.2010.5596842
- [17] Zhang, Z., Tao, D. (2012). Slow feature analysis for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(3): 436-450. https://doi.org/10.1109/TPAMI.2011.157
- [18] Chen, T.T., Ruan, Q.Q., An, G.Y. (2015). Slow feature extraction algorithm of human actions in video. CAAI Transactions on Intelligent Systems, 2015(3): 381-386.
- [19] He, H.H., Li, G.H., Yao, Q.S. (2014). Blind source separation of underwater acoustic signals by using slowness feature analysis. Technical Acoustics, 2014(3): 270-274. https://doi.org/10.3969/j.issn1000-3630.2014.03.017
- [20] Huang, J., Ersoy, O.K., Yan, X. (2017). Slow feature analysis based on online feature reordering and feature selection for dynamic chemical process monitoring. Chemometrics and Intelligent Laboratory Systems, 169: 1-11. https://doi.org/10.1016/j.chemolab.2017.07.013
- [21] Nie, W., Liu, A., Su, Y., Wei, S. (2017). Multi-view feature extraction based on slow feature analysis. Neurocomputing, 252: 49-57. https://doi.org/10.1016/j.neucom.2016.01.125
- [22] Hao, T., Wang, Q., Wu, D., Sun, J.S. (2018). Multiple person tracking based on slow feature analysis. Multimedia Tools and Applications, 77(3): 3623-3637.
- Yao, X.Z., Lu, T.W., Hu, H.P. (2009). Object recognition models based on primate visual cortices: A review. Pattern Recognition and Artificial Intelligence, 22(4): 581-588. https://doi.org/10.3969/j.issn.1003-6059.2009.04.012
- [24] Riesenhuber, M., Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11): 1019-1025. https://doi.org/10.1038/14819
- [25] Mutch, J., Lowe, D.G. (2006). Multiclass object recognition with sparse, localized features. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 1: 11-18. https://doi.org/10.1109/CVPR.2006.200
- [26] Avant, T., Morgansen, K.A. (2020). Analytical bounds on the local Lipschitz constants of affine-ReLU functions. arXiv preprint arXiv:2008.06141.
- [27] Zhang, W.L., Li, X.W., Song, Q.X., Lu, W. (2020). A face detection method based on image processing and improved adaptive boosting algorithm. Traitement du Signal, 37(3): 395-403. https://doi.org/10.18280/ts.370306
- [28] Ouali, M.A., Ghanai, M., Chafaa, K. (2020). TLBO optimization algorithm based-type2 fuzzy adaptive filter for ECG signals denoising. Traitement du Signal, 37(4): 541-553. https://doi.org/10.18280/ts.370401

- [29] Chen, T., Lasserre, J.B., Magron, V., Pauwels, E. (2020). Semialgebraic optimization for lipschitz constants of relu networks. arXiv preprint arXiv:2002.03657.
- [30] Latorre, F., Rolland, P., Cevher, V. (2020). Lipschitz

constant estimation of neural networks via sparse polynomial optimization. arXiv preprint arXiv:2004.08688.