

## Design and Application of a Slow Feature Algorithm Coupling Visual Selectivity and Multiple Long Short-Term Memory Networks



Yanming Zhao, Hong Yang\*, Guoan Su

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

Corresponding Author Email: [yanghonghot@sohu.com](mailto:yanghonghot@sohu.com)

<https://doi.org/10.18280/ts.380529>

### ABSTRACT

**Received:** 10 June 2021

**Accepted:** 28 August 2021

#### Keywords:

*slow features, long short-term memory network (LSTM), Lipschitz condition, visual selectivity, Gabor, visual computing*

In the traditional slow feature analysis (SFA), the expansion of polynomial basis function lacks the support of visual computing theories for primates, and cannot learn the uniform, continuous long short-term features through selective visual mechanism. To solve the defects, this paper designs and implements a slow feature algorithm coupling visual selectivity and multiple long short-term memory networks (LSTMs). Inspired by the visual invariance theory of natural images, this paper replaces the principal component analysis (PCA) of traditional SFA algorithm with myTICA (TICA: topologically independent component analysis) to extract image invariant Gabor basis functions, and initialize the space and series of basis functions. In view of the ability of the LSTM to learn long and short-term features, four LSTM algorithms were constructed to separately predict the long and short-term visual selectivity features of Gabor basis functions from the basis function series, and combine the functions into a new basis function, thereby solving the defect of polynomial prediction algorithms. In addition, a Lipschitz consistency condition was designed, and used to develop an approximate orthogonal pruning technique, which optimizes the prediction basis functions, and constructs a hyper-complete space for the basis function. The performance of our algorithm was evaluated by three metrics and mySFA's classification method. The experimental results show that our algorithm achieved a good prediction effect on INRIA Holidays dataset, and outshined SFA, graph-based SFA (SFA), TICA, and myTICA in accuracy and feasibility; when the threshold was 6, the recognition rate of our algorithm was 99.98%, and the false accept rate (FAR) and false reject rate (FRR) were both smaller than 0.02%, indicating the strong classification ability of our approach.

## 1. INTRODUCTION

Slow varying signals, an expression of the invariant, convey the invariant information of the high-level abstract expression of high-frequency input signals. Based on invariant learning, slow feature analysis (SFA) aims to extract the slow attributes of signals, as well as the slow topology between these attributes. The relevant algorithms have been successfully applied in many fields.

In 1989, Hinton [1] became the first to propose the basic concept, fundamental theories, and hypotheses of slow features, and formulated the preliminary theory of slow feature extraction. In 2002, Wiskott and Sejnowski [2] put forward an unsupervised learning algorithm for SFA, which overcomes the lack of samples and constant dimensionality of supervised learning approaches through nonlinear expansion of the feature space. This marks the official establishment of SFA theories and algorithms.

In 2005, Berkes and Wiskott [3] and Wiskott applied the invariant learning results to receptive field learning of complex cells. In 2007, Franzius et al. [4] revealed the properties of invariant on brain cells in the hippocampus, laying the physiological basis for invariant learning. In 2008, Franzius et al. [5] extracted and recognized the invariant features of the position and rotation angle of cartoon fish through SFA, and demonstrated the ability of SFA to extract classification information. Their study provides the theoretical

and experimental foundation for applying slow feature algorithms in feature extraction and pattern recognition.

In 2009, Kalmpfl and Maass [6] trained the SFA algorithm with the data generated by Markov chain, and proved that the features extracted by the slow feature algorithm are equivalent to those extracted by feature decay algorithm (FDA), when the small parameter  $\alpha$  approximates zero. The results show that invariant feature extraction algorithm is a correct and feasible way to extract features, and the Markov chain applies to slow feature training.

In 2011, Ma et al. [7] provided a kernel-based SFA algorithm, which enables the expansion of the feature space, prevents the computing in high-dimensional space, and applies SFA algorithm in blind signal separation. At this point, the standard SFA finally matured.

In 2012, Luciw et al. [8] presented a hierarchical SFA algorithm to reduce the computing complexity of the standard SFA. Inspired by the divide-and-conquer strategy, their algorithm can ease the computing complexity. But much information may get lost due to the hierarchical structure. Thus, the algorithm does not necessarily extract all global optimal features. In 2013, Escalante-B and Wiskott [9] developed an SFA based on the graph theory (GSFA) to overcome the difficulty in extracting global optimal features, which stems from information loss. The GSFA trains the SFA algorithm with the complex structure of the training images, trying to realize feature extraction according to the complex structure

features of the training images.

In 2015, Zhao et al. [10] designed a feature extraction algorithm based on the complex visual information of natural images. From the angle of visual selectivity, the self-designed topologically independent component analysis (TICA) was adopted to extract the slow features of the topological relationship between information in the complex visual space of natural images, revealing the consistency between visual selectivity and slow features.

In 2019, Zhao [11] improved the slow feature algorithm based on visual invariance, and successfully applied the improved algorithm to extract natural image features. In 2020, Zhao [12] developed an improved slow feature algorithm based on visual selectivity, and introduced it successfully to extract features from out-of-focus image series.

In 2020, Zhao [13, 14] improved the SFA algorithm by multiple long short-term memory networks (LSTMs) and spatiotemporal correlations, implemented the algorithm to dynamic forecast of particulate matter 25 (PM25), and demonstrated the excellent prediction ability of the LSTM-improved algorithm.

In recent years, the SFA algorithm has achieved desired effects in various fields, ranging from human behavior recognition [14-16], blind signal analysis [7, 17], dynamic monitoring [18], three-dimensional (3D) feature extraction [19], and multi-person path planning [20]. Great progress has been realized on the theories and applications of SFA. However, there remain two defects with traditional SFA: (1) The expansion of polynomial basis function lacks the support of visual computing theories for primates; (2) The expansion of polynomial basis function cannot learn the uniform, continuous long short-term features through selective visual mechanism. To solve the defects, this paper designs and implements a slow feature algorithm coupling visual selectivity and multiple LSTMs.

## 2. UNIFORM CONTINUITY OF VISUAL COMPUTING THEORIES AND ITS APPLICATION IN SLOW FEATURE ALGORITHM

Visual sparsity theory shows that the primates can express the unlimited nature with a limited number of visual neurons, and the expression can be illustrated with families of Gabor basis functions. According to the visual selectivity theory, the families of Gabor basis functions obey the visual selectivity rules of small intra-class distance and large between-class distance. Statistically, the basis functions of the same family have a low dispersion, while those of different families have a high dispersion. On image appearance, the Gabor basis functions containing similar visual selectivity features characterize the image blocks with similar textures, while those containing different visual selectivity features characterize image blocks with different textures. On the same image, the number of Gabor basis function families, and the basis functions in each family have stable visual information, and stable spatial topology of the information.

Under the constraint of visual selectivity theory, the Gabor basis functions in each family have stable and uniform global visual features, i.e., each Gabor basis function has uniform, continuous visual features. The LSTM algorithm can learn the

uniform, continuous property of the family (series) of basis functions effectively, and generate expanded Gabor basis functions through prediction. The visual selectivity consistency of the generated basis functions can be constrained by the Lipschitz uniform consistency theory. By this theory, it is possible to prune the basis functions, optimize the family of Gabor basis functions, and ensure the uniform, consistency of visual features in the function family.

According to the visual flow feature theory, when the primate visual system observes the external environment, the slow changes are covered up by the rapid changes of the observation features. In terms of images, the image signals contain a stable number and families of Gabor basis functions. In the same family, the visual features of Gabor basis functions are uniform and continuous. In the visual space, the Gabor basis functions in the same family have a slowly changing topology. The visual processing of primates is a layered serial process involving multiple parallel links [21, 22]. Through long-term training on an unlimited number of images, the links slowly form a stable structure and uniform, continuous functions. The link structure can learn the long and short-term visual selectivity features of images. The LSTM, as an improved recurrent neural network (RNN), can effectively learn or predict the long and short-term memory features of time series, thanks to its stable link structure. On visual information processing, the learning/prediction process bears resemblance to visual link theory. After link learning, the family of Gabor basis functions and their visual space topology become consistent in slow changing and visual selectivity.

Inspired by the theory of visual selection consistency, this paper adopts the multi-LSTM algorithm to predict the four key parameters of Gabor basis functions, and derives Gabor basis functions based on visual selectivity theory. This derivation approach was used to replace the basis function prediction method of SFA, i.e., the polynomial product method. This is the major innovative point of this research.

## 3. OUR ALGORITHM

### 3.1 Related results

The Introduction has already summarized the theories and applications of SFA, and identified the strengths and weaknesses of traditional SFA algorithms. Table 1 compares the research results related to our study.

Based on the comparison of Table 1, this paper pursues innovation in terms of slow feature extraction, flow feature basis function expansion, and feature structure. The visual selectivity theory was integrated with the Lipschitz uniform consistency theory, and LSTM theory. Specifically, a spatial expansion algorithm was developed for the basis function of the LSTM algorithm based on visual selectivity. The expansion method for slow feature basis functions was effectively integrated with the uniform, consistent visual selectivity, based on the approximate pruning of Lipschitz condition. On this basis, the hyper-complete set of out-of-focus image series was predicted and extracted, supporting the recognition of the out-of-focus image series. The innovation process of our algorithm is introduced step by step below.

**Table 1.** Comparison of relevant results

Contents	Key literature	Feature extraction method	Feature expansion method	Feature structure
Slow features	Wiskott and Sejnowski in 2002 [2]	PCA	Polynomial expansion	Global set of slow features
	Luciw et al. in 2012 [8]	PCA	Polynomial expansion	Slow feature set of hierarchical sampling
	Escalante-B. and Wiskott in 2013 [9]	PCA	Polynomial expansion	Graph structure feature set
	Zhao in 2015 [10]	myTICA	Feature-based probability distribution	Slow feature set of visual space topology
	Zhao in 2019-2020 [11, 12]	mySFA	Visual invariance and visual selectivity	3D visual information and information of visual space topology
Improvement of LSTM prediction algorithm	Zhao in 2020 [13]	Spatiotemporal correlated micro-features of long and short-term memory	Not involved	Global and local spatial correlation matrices with long and short-term properties (microscale)
	Zhao in 2020 [14]	Spatiotemporal correlated macro-features of long and short-term memory	Not involved	Global and local spatial correlation matrices with long and short-term properties (macroscale)

### 3.2 Innovations

To overcome the defects of the original SFA, our algorithm achieves innovation in the following four aspects, based on the above theoretical analysis.

Innovation 1: receptive field modeling based on visual selectivity

The receptive field of visual cells can be modeled by Gabor function, which uses four key parameters to depict the visual selectivity of the visual cells of primates. The four parameters contribute differently to selectivity. By the degree of contribution, they are ranked as frequency, direction, phase, and visual space position. Based on the theory of visual selectivity, the generalized Gabor model can be established as:

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

where,  $\lambda_f$  and  $\lambda_o$  are the frequency and direction of visual space in the receptive field of visual cells, respectively (as distinguishing properties of the receptive field of visual cells, these two parameters determine the visual space distribution, and bear on the calculation of intra-class and between-class dispersions);  $\lambda_p$  is the visual spatial lag of the receptive field of primary visual cells (this parameter plays an important role in the calculation of intra-class dispersion in the receptive field of primary visual cells);  $\lambda_d$  is the position of the receptive field of primary visual cells in the visual space (this parameter plays an important role in the calculation of between-class dispersion). The four parameters work together to determine the global properties of the receptive field.

Inspired by visual selectivity theory, the four parameters of the Gabor function can be normalized by contribution:

$$\bar{\lambda}_i = \frac{|\lambda_i|}{\sqrt{\lambda_f^2 + \lambda_o^2 + \lambda_p^2 + \lambda_d^2}} \quad (2)$$

Experimental results show that the proportional relationship of the four parameters of the receptive field was  $\bar{\lambda}_f : \bar{\lambda}_o : \bar{\lambda}_p : \bar{\lambda}_d = 4 : 4 : 1 : 1$ , which basically reflects the degree and amount of contribution of each receptive field parameter.

Innovation 2. Improvement of feature extraction method

In the original SFA algorithm, the features are extracted through PCA. The main components are mined from the original image, while the visual space topology of the natural image is not extracted. Thus, the PCA results cannot convey the intrinsic visual features of the image. Inspired by the theory of visual selectivity, the basis functions in the basis function space  $Visual\_sub\_set(i)$  have similar or the same global visual selectivity, and are distributed close in the visual space. Therefore, these functions are clearly uniform and consistent in visual selectivity. Hence, the PCA in the original SFA algorithm was replaced with myTICA [20] to extract the slow visual features from the natural image, that is,  $Visual\_sub\_set$  was modeled by Gabor function as:

$$\begin{aligned} Visual\_sub\_set &= \{Visual\_sub\_set\{i\} \\ &= Gabor(\lambda_f, \lambda_o, \lambda_p, \lambda_d) | i = 1, 2, 3, \dots, N\} \end{aligned} \quad (3)$$

where, f, o, p, and d are the frequency, direction angle, phase, and visual space location of the receptive field, respectively.

Innovation 3. The spatial expansion algorithm for the basis functions of the LSTM algorithm based on visual selectivity (Base\_Extend\_of\_LSTM\_Selectivity, BELS).

Originally, the feature space of the SFA algorithm is mostly expanded by polynomial method. The complex computing process lacks the support from visual computing theories for primates, or the ability to learn the long and short-term features, which are uniform and continuous in visual selectivity. From the intrinsic visual features, this paper expands the feature map to generate a hyper-complete set of invariant features, which reflects the nature of natural images.

### 3.3 Algorithm structure

The flow of the BELS algorithm is shown in Figure 1.

The basis function space is formed by dividing the basis functions of the natural image. Reflecting the visual space selectivity of the natural image, the basis function space can be expressed as  $Visual\_sub\_set = \{Visual\_sub\_set\{i\} | i = 1, 2, 3, \dots, N\}$ . Inspired by the visual selectivity of primates (small intra-class distance and large between-class distance), the basis function space  $Visual\_sub\_set(i)$  is divided into four subsets  $Visual\_sub\_set\{k, \lambda 1\}$ ,  $Visual\_sub\_set\{k, \lambda 2\}$ ,  $Visual\_sub\_set\{k, \lambda 3\}$  and  $Visual\_sub\_set\{k, \lambda 4\}$  by the degree

of contribution of basis function parameters. The technical pointer of the dynamics of basis function space is described by a counter  $p\_add$ , while the number of elements in the basis function space is depicted as number1. The K-subset generation algorithm collects K samples randomly from the set

$Visual\_sub\_set(i)$ . The value of K is smaller than or equal to the number of elements in that set. The LSTM operations in Gabor frequency, direction, phase and spatial distance are formulated into four algorithms: F\_LSTM, O\_LSTM, P\_LSTM and D\_LSTM.

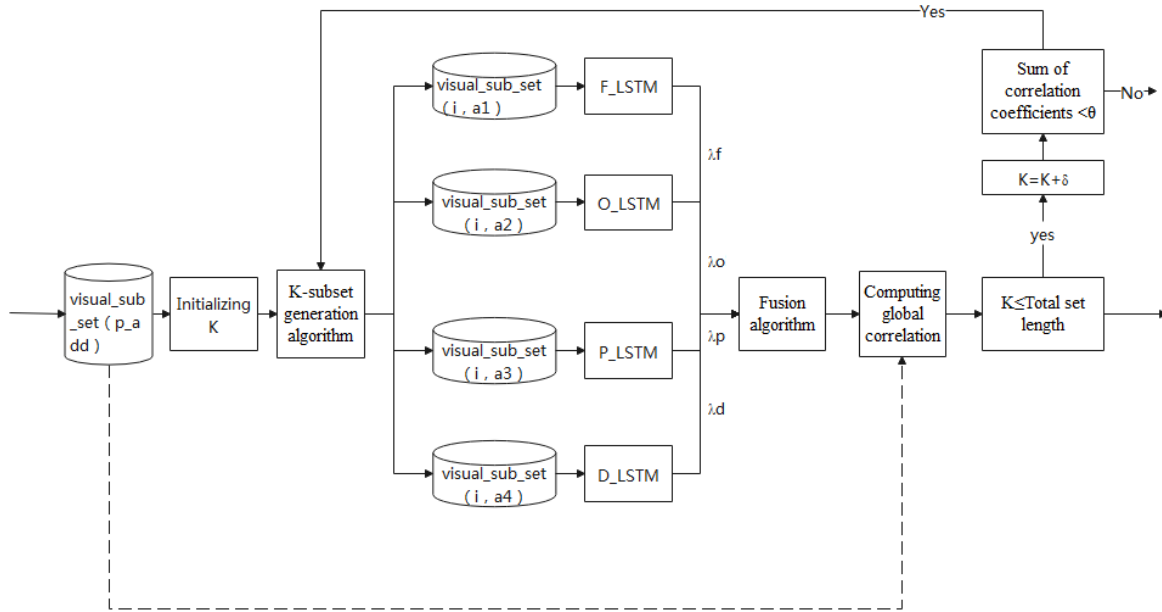


Figure 1. Flow of BELS algorithm

### 3.4 BELS algorithm

Step 1. Initialize  $k = get\_length(Visual\_sub\_set\{p\_add\})/2$ , and extract subset  $Visual\_sub\_set\{p\_add, k\}$  from the normalized  $Visual\_sub\_set$ , where  $get\_length(*)$  is the number of elements in set \*. The K value is initialized as half of the length of  $Visual\_sub\_set\{p\_add\}$ . The K randomly selected elements from  $Visual\_sub\_set$  are grouped into  $Visual\_sub\_set\{p\_add, k\}$ .

Step 2. Using K-subset generation algorithm to randomly choose k elements from normalized  $Visual\_sub\_set$ , and group them into a subset  $Visual\_sub\_set\{p\_add, k\}$ . Then, sequentially import  $Visual\_sub\_set\{p\_add, k, \lambda_1\}$ ,  $Visual\_sub\_set\{p\_add, k, \lambda_2\}$ ,  $Visual\_sub\_set\{p\_add, k, \lambda_3\}$ , and  $Visual\_sub\_set\{p\_add, k, \lambda_4\}$  to algorithms F\_LSTM, O\_LSTM, P\_LSTM, and D\_LSTM, thereby predicting parameters  $\lambda_1, \lambda_2, \lambda_3$ , and  $\lambda_4$  and obtaining visual selectivity parameters  $\lambda_f, \lambda_o, \lambda_p$ , and  $\lambda_d$ . The visual selectivity parameters are thereby predicted.

Step 3. Normalize parameters based on visual selectivity, using contribution-based normalization method (2). In this way, the normalized parameters  $\lambda_f, \lambda_o, \lambda_p$ , and  $\lambda_d$  are obtained.

Step 4. Output the predicted parameter vector  $\alpha = (\lambda_f, \lambda_o, \lambda_p, \lambda_d)$  of Gabor basis functions, and normalize the relevant count  $relation\_s = 0$ . Given that  $\forall \beta = (\bar{\lambda}_f, \bar{\lambda}_o, \bar{\lambda}_p, \bar{\lambda}_d) \in Visual\_sub\_set(p\_add)$ , compute  $Correlation\_value = Correlating\_operation(\alpha, \beta)$ . If  $Correlation\_value \geq \gamma$ , then  $relation\_s ++$ , where  $\gamma = 0.8$  is a preset threshold for correlation. Traverse the entire  $Visual\_sub\_set(p\_add)$  to compute the correlation coefficient. If  $relation\_s \leq \theta, \theta = \rho \times len(Visual\_sub\_set(p\_add))$ , vector  $\alpha$  is not related to set  $Visual\_sub\_set(p\_add)$ . Then, K should be increased by a step length  $\delta$ , i.e.,  $k = k + \delta$  and  $k \leq$

$len(Visual\_sub\_set(p\_add))$ , with  $\delta$  be a parameter controlling the step length, which is measured through experiments. After that, jump to Step 2. Otherwise, go to Step 5.

Step 5. Output prediction parameter  $\alpha = (\lambda_f, \lambda_o, \lambda_p, \lambda_d)$  to  $Visual\_sub\_set(p\_add)$ .

### 3.5 Approximate orthogonal pruning based on Lipschitz condition (Lipschitz\_Orthogonal\_Pruning\_Method, LOPM)

In series  $Visual\_sub\_set = \{\gamma(i) | \gamma(i) = (\lambda_f^i, \lambda_o^i, \lambda_p^i, \lambda_d^i), i=1,2,3,\dots,M,\dots,N\}$ , the first M elements are basis functions extracted by myTICA, and the elements M+1 to N are the expanded basis functions predicted by our algorithm. Let  $(\mu, \Sigma)$  be the mean and covariance of samples, respectively;  $\forall \beta, \bar{\beta} \in Visual\_sub\_set$ ;  $\alpha = \{\lambda_f^*, \lambda_o^*, \lambda_p^*, \lambda_d^*\}$  be the basis function parameter vector predicted by the LSTM algorithm.

This paper proposes the LOPM to ensure that the predicted basis function  $\alpha$  and the functions in the set  $Visual\_sub\_set$  have similar, uniform, and continuous visual features.

According to Lipschitz consistency [23-25], the following inequalities can be obtained:

$$\begin{cases} D(Gabor(\alpha), Gabor(\beta)) \leq L \times D(\alpha, \beta) \\ \alpha \leq Max\{\mu + \alpha \Sigma\} \end{cases}$$

Since

$$D(\alpha, \beta) \leq Max(D(\alpha, \beta), D(\beta, \bar{\beta}))$$

When  $\bar{L} = Max(D(\alpha, \beta), D(\beta, \bar{\beta}))$ , the finetuning parameter of consistency  $\eta = \frac{D(\alpha, \beta)}{D(\alpha, \beta) + D(\beta, \bar{\beta})}$ . If,

$$L = \begin{cases} \bar{L} & \eta \geq \xi \\ \bar{L} + \eta & \eta < \xi \end{cases}$$

where,  $\xi$  is a random small integer preventing overconvergence of the elements in the set  $Visual\_sub\_set$ ;  $L \ll D(Visual\_sub\_set(k1), Visual\_sub\_set(K2))$ ;  $k1 \neq K2, k1, k2 \leq length(Visual\_sub\_set)$ .

Then, there must be:

$$D(Gabor(\alpha), Gabor(\beta)) \leq L \times D(\alpha, \beta)$$

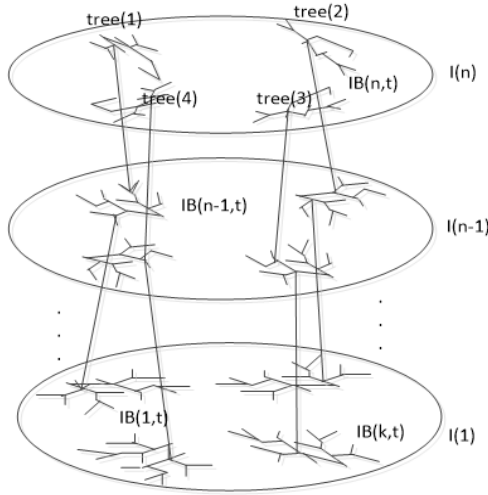
Hence, the value of L is determined, and the Lipschitz consistency is basically guaranteed. In this case, function D is introduced to measure the block distance between two vectors:

$$D(\alpha, \beta) = |\lambda_f^\alpha - \lambda_f^\beta| + |\lambda_o^\alpha - \lambda_o^\beta| + |\lambda_p^\alpha - \lambda_p^\beta| + |\lambda_d^\alpha - \lambda_d^\beta|$$

If  $\langle \alpha, \beta \rangle \leq \varepsilon$ , vector  $\alpha$  is similar to vector  $\beta$ , carry out pruning and retain  $\beta$ ; otherwise, add  $\alpha$  to set  $Gabor\_set$ . Note that  $\varepsilon$  is pruning coefficient, a random small integer controlling the degree of pruning.

#### Innovation 4. Invariant feature formation

The original SFA algorithm extracts only the PCA features of the natural image, failing to extract slow features from out-of-focus image series or the spatial topology between these features, using the visual computing theory of natural images.



**Figure 2.** Feature forest reflecting the visual invariance of the natural out-of-focus image series

Note:  $tree(m), m=1,2,3,\dots$  is the  $m$ -th series feature of the out-of-focus image series in the visual space. It is designed according to the global parallel and local serial processing mechanism of visual information of primates, with the aim to facilitate parallel algorithm design.

Inspired by the theory on the consistency of visual selectivity of primates, it is certain that the Gabor basis functions satisfy the consistency of visual selectivity in the complex visual space  $Visual\_sub\_set$ , which corresponds to the target out-of-focus image series. Due to the uniformity and continuity (or invariance) of the four visual parameters  $f, o, p,$  and  $d$  of basis functions determine the number of basis function spaces and the spatial topological relationship of the out-of-focus image series, as well as the consistency between the type of basis functions in the basis function family and

spatial topology.

Through the above improvement, the feature forest can be constructed to reflect the visual invariance of the out-of-focus image series, based on the improved SFA architecture (Figure 2).

To sum up, the slow features of the out-of-focus image series can be expressed as:

$$IS = \{tree(m) | m=1,2,3,\dots\} \quad (5)$$

$$tree(m) = \{(index(m, j), gabor(m, j))\} \quad (6)$$

where,  $m$  is the visual space basis;  $j$  is the position of an element in  $m, j=1,2,3,\dots,360$ . In the IS matrix, the columns have frequency selectivity, and the rows have direction angle selectivity. The elements are ranked in ascending order, with the direction angle falling in  $[1^\circ, 360^\circ]$ . Hence, the features can be described by the following matrix.

**Table 2.** Feature matrix  $index$  (each row contains 360 features)

$Index(l,j)$	1	0	0	...	1
$Index(2,j)$	0	0	1	...	0
$Index(3,j)$	1	1	0	...	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Index(n,j)$	0	1	1	...	1

In the feature matrix  $index$ , 1 means the receptive field exists at that position; otherwise, the receptive field does not exist at that position. Each node  $index(i,j)$  contains the corresponding  $gabor(i,j)$  function. These functions form a feature matrix (Tables 2 and 3).

**Table 3.** Receptive fields corresponding to GMindex

$chain(1)$	$g11$	$g12$	$g13$	...	$g1360$
$chain(2)$	$g11$	$g12$	$g13$	...	$g1360$
$chain(3)$	$g11$	$g12$	$g13$	...	$g1360$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$chain(n)$	$g11$	$g12$	$g13$	...	$g1360$

## 4. ALGORITHM DESIGN

### 4.1 Algorithm structure

Reflecting the above innovations, the proposed basis function expansion algorithm can be implemented in the following flow (Figure 3).

The basis function space is formed by dividing the basis functions of the natural image. Reflecting the visual space selectivity of the natural image, the basis function space can be expressed as  $Visual\_sub\_set = \{Visual\_sub\_set\{i\} | i, 2, 3, \dots, N\}$ . Inspired by the visual selectivity of primates (small intra-class distance and large between-class distance), the basis function space  $Visual\_sub\_set(i)$  is divided into four subsets  $Visual\_sub\_set\{k,\lambda1\}, Visual\_sub\_set\{k,\lambda2\}, Visual\_sub\_set\{k,\lambda3\}$  and  $Visual\_sub\_set\{k,\lambda4\}$  by the degree of contribution of basis function parameters. The technical pointer of the dynamics of basis function space is described by a counter  $p\_add$ , while the number of elements in the basis function space is depicted as number1. The K-subset

generation algorithm collects K samples randomly from the set  $Visual\_sub\_set(i)$ . The value of K is smaller than or equal to the number of elements in that set. The LSTM operations in

Gabor frequency, direction, phase and spatial distance are formulated into four algorithms: F\_LSTM, O\_LSTM, P\_LSTM and D\_LSTM.

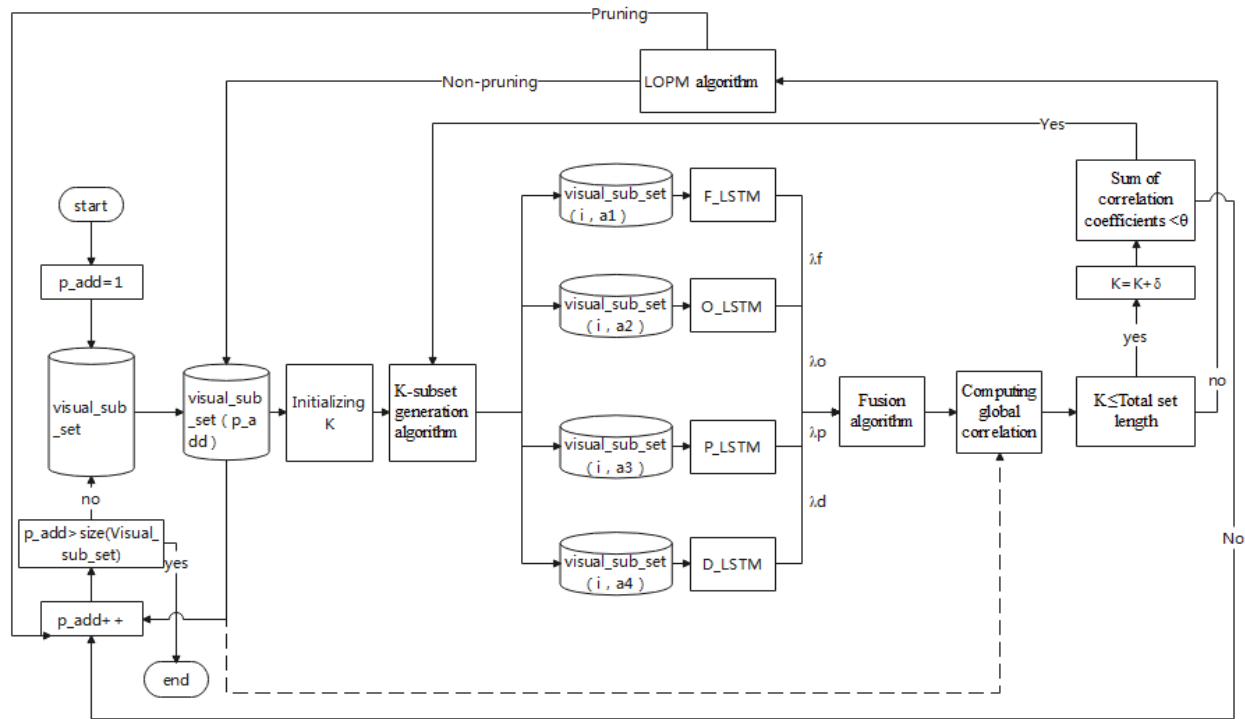


Figure 3. Flow of our algorithm

#### 4.2 Feature prediction algorithm

Step 0. Initialize control variable  $p\_add=1$ .

Step 1. Establish the set of description information for primates' visual space by myTICA [10]:  $Visual\_sub\_set = \{Visual\_sub\_set(p\_add)|i, 2, 3, \dots, N\}$ , and compute  $number1 = (operation(Visual\_sub\_set\{p\_add\}))$ .

Step 2. Normalize parameters in each subset of  $Visual\_sub\_set$  based on visual selectivity, using contribution-based normalization method (2).

Step 3. Call the BELS algorithm to obtain the prediction vector  $\alpha = (\lambda_f, \lambda_o, \lambda_p, \lambda_d)$ .

Step 4. Call the LOPM. If pruning is executed, jump to Step 5; otherwise, insert vector  $\alpha = (\lambda_f, \lambda_o, \lambda_p, \lambda_d)$  into  $Visual\_sub\_set(p\_add)$  before jumping to Step 5.

Step 5.  $p\_add++$ .

Step 6. If  $p\_add \leq number1$ , jump to Step 3.

Step 7. Otherwise, terminate the algorithm, and complete the prediction of slow features.

Step 8. Generate the feature forest  $IS = \{tree(m) | m = 1, 2, 3, \dots\}$ .

#### 4.3 Prediction evaluation and classification algorithm

The algorithm performance is evaluated by three metrics: root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE):

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

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

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

where,  $y_i^*$  is the basis function generated by myTICA;  $y_i$  is the basis function predicted by our algorithm;  $n$  is the number of observations. The classification is carried out by Zhao's method [11].

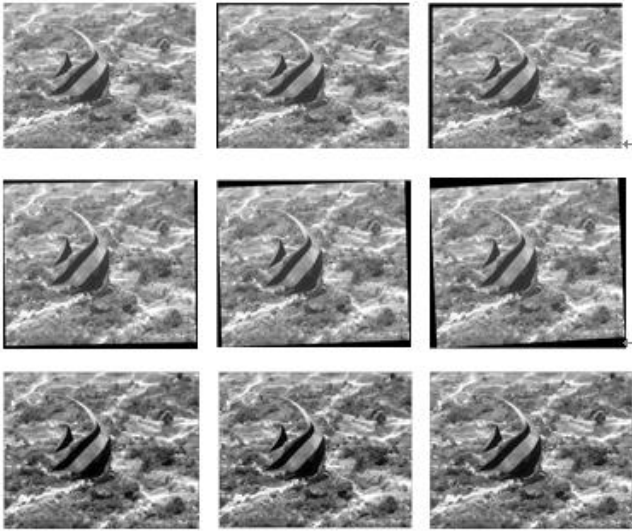
### 5. EXPERIMENTS AND RESULTS ANALYSIS

#### 5.1 Image set generation and display

Table 4. Affine transform parameters (the last column is for comparison)

Type	Parameters (rotation angle, number of translation pixels, and scaling factor)					
Rotation	0.1	0.3	0.5	0.7	1	<u>5</u>
Translation	1	3	5	7	9	<u>9</u>
Scaling	1/8	1/4	1/2	2	3	<u>4</u>

According to the previous theoretical analysis, out-of-focus images can be approximated through the affine transform of original clear images. Hence, 1,000 images are selected from INRIA Holidays dataset, and subjected to affine transform using the parameters in the first five columns of Table 4, producing an image set of 15,000 images. The first 12,000 images were allocated to the training set, and the latter 3,000 into the test set. Each set contains 1,000 original images. Figure 4 shows the samples from the experimental image set.



**Figure 4.** Samples from the experimental image set based on INRIA holidays dataset

### 5.2 Relationship between initial K and BELS performance

In our algorithm, the K value determines the visual selectivity consistency of the predicted basis functions. It also bears on the prediction and computing performance of the BELS. This parameter could be configured adaptively according to the learning results. Therefore, the initial K value determines the prediction performance and execution process of the BELS. The selection of K value is critical to whether the predicted basis functions can learn the global distribution features of set  $Visual\_sub\_set(i)$ . Experiments show that the initial K should satisfy  $K = \text{Min}(\text{length}(\text{visual\_sub\_set}(i)), i = 1, 2, 3, \dots, N)/2$ . Through random sampling, the K-subset generation algorithm collects K elements from set  $visual\_sub\_set(i)$ , forming a new set  $Visual\_sub\_set(i, \lambda_i)$  with  $t \in (\lambda_f, \lambda_o, \lambda_p, \lambda_d)$ .

On the training set, the relationship of initial K and the prediction performance of the BELS is shown in Table 5.

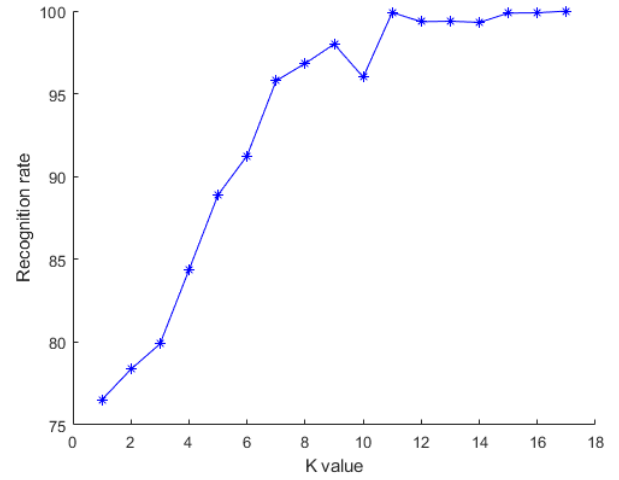
With the growth of initial K, more basis functions in  $Visual\_sub\_set(i)$  participate in the prediction, and the learnable global visual features are enhanced. Meanwhile, the predicted basis function  $\alpha$  has stronger global features of the basis function set  $Visual\_sub\_set(i)$ , and exhibit a more uniform and continuous visual selectivity. The results of RMSE, MAE, and MAPE all gradually decrease. Therefore, the BELS can achieve a very strong prediction performance.

**Table 5.** Relationship of initial K and our algorithm's prediction performance

Metric \ K	N/2	$\frac{7}{16}N$	$\frac{5}{8}N$	$\frac{11}{16}N$	$\frac{3}{4}N$	$\frac{13}{16}N$	$\frac{7}{8}N$	$\frac{15}{16}N$	N
RMSE	15.23	14.78	13.82	13.15	12.09	11.54	9.78	9.02	8.36
MAE	7.91	7.53	7.05	6.85	6.28	6.08	5.89	5.56	5.31
MAPE	9.82	9.45	9.03	8.86	8.19	7.56	7.12	6.64	6.01

**Table 6.** Relationship between parameter  $\delta$  and our algorithm's prediction performance

Metric \ $\delta$	-5	-4	-3	-2	-1	1	2	3	4	5
RMSE	15.67	14.54	13.61	12.45	11.63	10.75	9.01	8.36	7.26	6.76
MAE	8.98	8.32	7.86	7.01	6.52	5.76	5.09	4.65	4.01	3.81
MAPE	12.16	11.82	11.09	10.76	9.67	8.54	7.69	6.54	5.84	4.98



**Figure 5.** Relationship between k value and BELS performance

The K value has the following correspondence with  $K$ :  $K(k\_value) = \frac{N}{2} + (k\_value - 1) \times d$ , where  $d=1/38$  is the tolerance;  $N$  is the size of  $Visual\_sub\_set(i)$ .

As shown in Figure 5, with the growth of initial K, the BELS witnesses an improvement in prediction performance; the predicted basis function  $\alpha$  can better learn and inherit the global features of elements in set  $Visual\_sub\_set(i)$ , and achieve a more uniform and continuous visual selectivity. Therefore, the recognition performance of our algorithm increases with the K value.

### 5.3 Relationship between initial $\delta$ and prediction performance

In our algorithm, the growth rate of K is described by parameter  $\delta$ :

$$\begin{cases} \delta = \delta_0 \\ \delta_t = \kappa_t - k_{t-1} \\ k_{t+1} = k_t + \delta_t \end{cases} \quad (10)$$

This parameter is an integer that determines the learning rate of the BELS. Hence, the initial value of this parameter is correlated with the prediction performance of the algorithm. The relationship between parameter  $\delta$  and parameter K on the training set is shown in Table 6.

As initial  $\delta$  increases from -5 to 5, the K value increases, and the predicted basis function  $\alpha$  can better learn and inherit the global features of elements in set  $Visual\_sub\_set(i)$ , and achieve a more uniform and continuous visual selectivity. That is, the BELS acquires stronger learning ability. Thus, the greater the K value, the smaller the changes of RMSE, MAE, and MAPE.

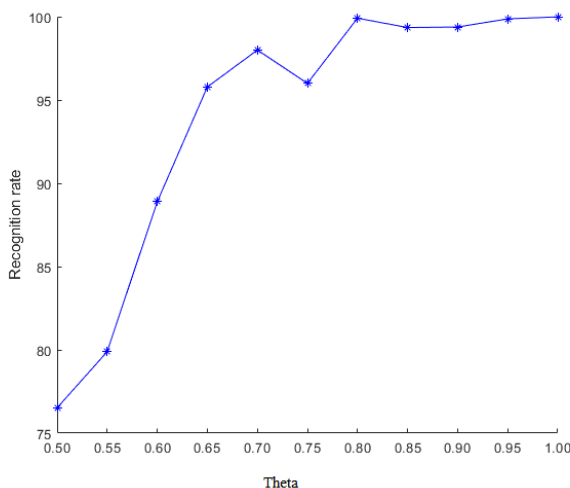
#### 5.4 Relationship between initial $\theta$ and prediction and recognition performance

In this paper, parameter  $\theta \in [0,1]$  describes the uniformity and continuousness of visual features between predicted basis function  $\alpha = \{\lambda_f^*, \lambda_o^*, \lambda_p^*, \lambda_d^*\}$  and basis function  $\beta \in Visual\_sub\_set(i)$ . Table 7 shows the relationship between  $\theta \geq 0.5$  and the prediction ability of our algorithm.

**Table 7.** Relationship between parameter  $\theta$  and our algorithm's prediction performance

Metric \ $\theta$	0.5	0.6	0.7	0.8	0.9	1
RMSE	10.75	10.03	9.54	9.05	8.76	8.45
MAE	5.76	5.10	4.52	4.22	3.92	3.27
MAPE	8.54	7.89	7.50	6.34	6.09	5.31

As parameter  $\theta$  increases from 0.5 to 0.8, the uniformity and continuousness of visual features improve between basis functions  $\alpha, \beta$ . The visual selectivity of predicted basis function  $\alpha$  are uniform and correlated with those of the basis function set  $Visual\_sub\_set(i)$ . The greater the value of  $\theta$ , the predicted basis function  $\alpha$  can better learn and inherit the global features of elements in set  $Visual\_sub\_set(i)$ , achieve a more uniform and continuous visual selectivity, and reduce the error with  $Visual\_sub\_set(i)$ . Thus, the greater the  $\theta$  value, the smaller the changes of RMSE, MAE, and MAPE.



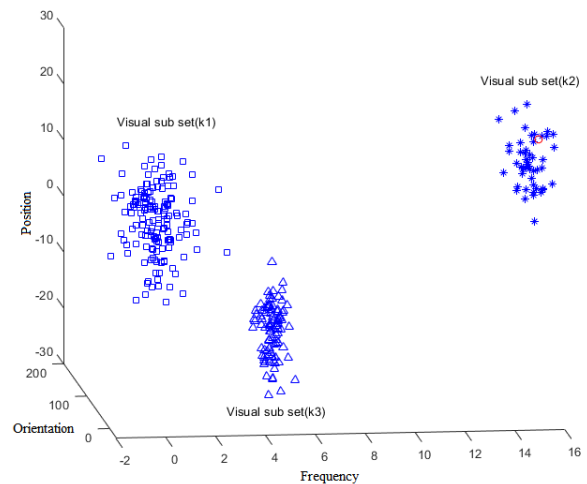
**Figure 6.** Relationship between parameter  $\theta$  and our algorithm's prediction performance

As shown in Figure 6, when parameter  $\theta$  increases from 0.5 to 0.8, the uniformity and continuousness of visual features improve between basis functions  $\alpha, \beta$ ; the predicted basis function  $\alpha$  can better learn and inherit the global features of elements in set  $Visual\_sub\_set(i)$ , achieve a more uniform and continuous visual selectivity, and acquire a stronger classification ability. Meanwhile, the algorithm becomes more

excellent in classification, at the cost of rising time complexity. Overall, our algorithm has the best prediction effect at  $\theta=0.8$ .

#### 5.5 Distribution consistency between predicted basis function and basis function set

Using the above optimal parameters, our algorithm is executed on the training set. The distribution of the predicted basis function  $\alpha = \{\lambda_f^*, \lambda_o^*, \lambda_p^*, \lambda_d^*\}$  and that of  $\forall \beta, \beta \in Visual\_sub\_set(i)$  are obtained as shown in Figure 7.



**Figure 7.** Distribution of the predicted basis function and basis function set

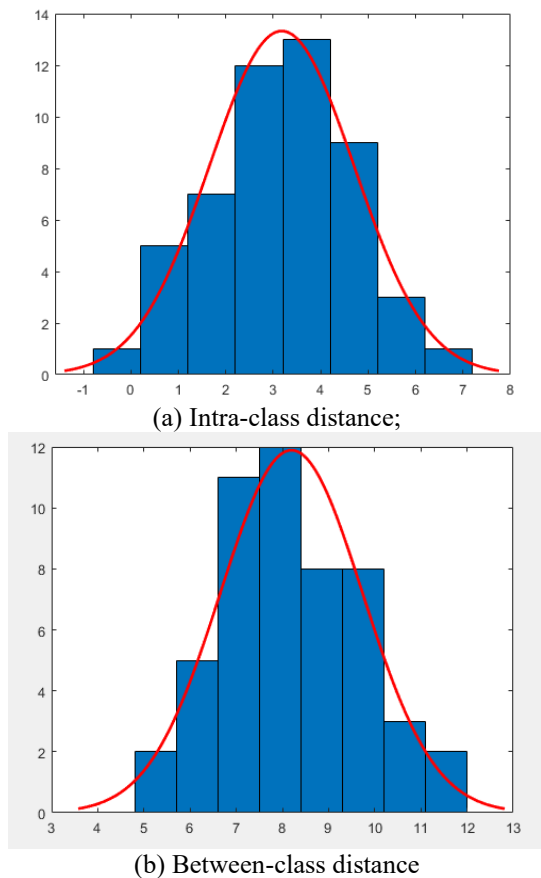
The global visual selectivity of basis function set  $Visual\_sub\_set(k3)$  is learned by the basis function  $\alpha$  predicted by the BELS. The predicted basis function  $\alpha$  achieves a strong consistency in visual selectivity with the set. In terms of visual selectivity, the predicted basis function has a small intra-class dispersion with its family  $Visual\_sub\_set(k2)$ , and a large between-class dispersion with the other families  $Visual\_sub\_set(k1)$  and  $Visual\_sub\_set(k3)$ . This satisfies the theoretical requirements of uniform and continuous visual selectivity within the class, and the differential visual selectivity across classes, and demonstrates the visual selectivity features of primates. Hence, our BELS can generate basis functions consistent in visual selectivity, and uniformly distributed across families.

#### 5.6 Feasibility of our algorithm

Using  $k=N/2$ ,  $\sigma=1$ , and  $\theta=0.75$ , our algorithm is applied to the training set. The intra-class and between-class distribution performance of our algorithm is recorded in Figure 8.

Our algorithm is improved from the original SFA algorithm with the aid of BELS and LOPM. In visual selectivity, the predicted basis function has a strong uniformity and consistency with the basis function set  $Visual\_sub\_set(k3)$ . Hence, the predicted basis function  $\alpha$  has a relatively small intra-class dispersion and a relatively large between-class dispersion. Under  $k=N/2$ ,  $\sigma=1$ ,  $\theta=0.75$ , and threshold of 6, the receiver operating characteristic (ROC) curve shows that the recognition rate of our algorithm was 99.98%, and the false accept rate (FAR) and false reject rate (FRR) were both smaller than 0.02%, indicating the strong classification ability of our approach.

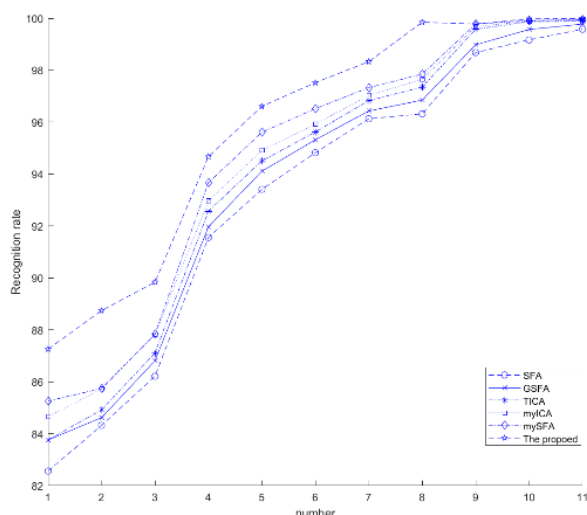




**Figure 8.** Intra-class and between-class distribution performance of our algorithm

### 5.7 Superiority of our algorithm

Using  $k=N/2$ ,  $\sigma=1$ , and  $\theta=0.75$ , our algorithm is compared with SFA [1], GSFA [8], TICA [9], myTICA [10], and mySFA [11] (Figure 9).



**Figure 9.** Recognition performance of different algorithms

Based on these conventional approaches, our algorithm replaces the polynomial method of basis function expansion in the original SFA with the BELS, such that the predicted basis function  $\alpha$  is highly uniform and continuous with the basis function set  $Visual\_sub\_set(i)$  in terms of visual selectivity. In addition, LOPM pruning technique is adopted to realize the

Lipschitz consistency condition of the predicted basis function. As a result, under  $k=N/2$ ,  $\sigma=1$ , and  $\theta=0.75$ , the basis function  $\alpha$  predicted by our algorithm has strong uniformity and continuity, and realizes uniform and continuous visual selectivity. Besides, the basis function boasts a small intra-class dispersion and a high between-class dispersion. These results show that our algorithm has a stronger classification effect than conventional approaches.

## 6. CONCLUSIONS

This paper designs and applies a slow feature algorithm that fuses visual selectivity and multi-LSTM algorithm. Specifically, myTICA is adopted to replace the PCA of conventional SFA, aiming to extract the Gabor basis functions reflecting the visual invariance of natural images, and to initialize the basis function space and basis function set. Drawing on the learning ability of the LSTM of long and short-term features, four LSTM algorithms are established to predict the long and short-term visual selectivity features of Gabor basis functions, based on the basis function set. The Gabor basis functions are combined into a new basis function, overcoming the defect of the polynomial prediction of conventional SFA. In addition, the predicted basis function and basis function space are optimized by designing a Lipschitz condition and proposing an approximation pruning technique under that condition. Through experiments on images obtained from INRIA Holidays dataset, our algorithm is found to have an excellent prediction performance. The algorithm is superior to and more feasible than SFA, GSFA, TICA, myTICA, and mySFA. At the threshold of 6, the recognition rate of our algorithm was 99.98%, and the FAR and FRR were both smaller than 0.02%, indicating the strong classification ability of our approach.

To further improve its application effect in visual computing, our algorithm will be further optimized in terms of parameter initialization of BELS, and the construction of Lipschitz condition. The future work will also explore the application of the algorithm in signal analysis and recognition.

## 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).

## 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>
- [2] 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*, pp. 961-970. [https://doi.org/10.1007/978-3-540-87536-9\\_98](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). Kernel-based 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. (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] 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>
- [13] 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>
- [14] Zhao, Y. (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] 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>
- [16] 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>
- [17] 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. <https://doi.org/10.3969/j.issn.1673-4785.201407002>
- [18] He, H.H., Li, G.H., Yao, Q.S., He, X.K., Shi, C.X. (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>
- [19] 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>
- [20] 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>
- [21] 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. <https://doi.org/10.1007/s11042-017-5218-4>
- [22] 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>
- [23] Riesenhuber, M., Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11): 1019-1025. <https://doi.org/10.1038/14819>
- [24] Liu, Y., Cao, W., Li, Y. (2022). Split-step balanced  $\theta$ -method for SDEs with non-globally Lipschitz continuous coefficients. *Applied Mathematics and Computation*, 413: 126437. <https://doi.org/10.1016/j.amc.2021.126437>
- [25] Fu, J., Ma, Z., Fu, Y., Chai, T. (2021). Hybrid adaptive control of nonlinear systems with non-Lipschitz nonlinearities. *Systems & Control Letters*, 156: 105012. <https://doi.org/10.1016/j.sysconle.2021.105012>
- [26] Li, X.N. (2011). The strong convergence of uniformly L-Lipschitzian mappings in Banach space. *Journal of University of Science and Technology of Suzhou (Natural Science)*, 28(1): 1-5. <https://doi.org/10.3969/j.issn.1672-0687.2011.01.001>