

CLASSIFICATION OF VIDEO DYNAMIC TEXTURE BASED ON LIE-SVM

Zhigui Li¹, Zhiyuan Zhu², Mingliang Zheng^{3✉}

Received on March 5, 2026

Presented by S. Hadjitodorov, Corresponding Member of BAS, on June 30, 2026

Abstract

In order to more accurately measure the dynamic texture pattern of video clips, it is necessary to analyze the topological space of this pattern; to this end, we propose a method for classifying video clips based on dynamic texture features and a support vector machine (SVM) with a Lie group kernel. First, the autoregressive moving average model (ARMA) is established to describe the dynamic texture of video, which is mapped to the Lie group manifold to form the matrix shape of Gaussians (SOG) descriptors. Second, the distance measurement criterion on Lie group manifold is used as the kernel function of SVM, and the algorithm of multi-classifier with Lie-SVM is designed. Finally, the example verifies that the multi-classifier for dynamic texture features of flame video based on Lie-SVM has better recognition rate (accuracy improvement 15.26%) and lower time consumption (runtime reduction 2.3%), which provides a new idea for the recognition of dynamic image texture features. The results show that the average recognition rate of flame video dynamic texture based on Lie-SVM algorithm is 11% higher than that of Martin distance algorithm, and through the influence of the dimension of state sequence on the average recognition rate, it can be concluded that the optimal condition of Lie-SVM algorithm is the singular value decomposition of observation sequence with 12 dimension.

Key words: video classification, dynamic texture, SOG descriptors, Lie-SVM

Introduction. Video automatic classification technology [1] is one of the most important and challenging research hotspots in the field of computer vision. At present, there are three kinds of video classification: artificial video annotation classification [2], active learning video annotation classification [3], and machine learning video annotation classification [4]. Among them, the machine learning is a very efficient video classification technology because it can adopt the supervised, semi-supervised and unsupervised flexible learning methods for small sample set. Support Vector Machine (SVM) [5] is a representative statistical machine learning algorithm proposed by BOSER et al. in 1992. It overcomes some problems of traditional classification methods, such as model selection, over learning, non-linear, multi-dimensional, and local minima [6, 7]. Therefore, video recognition based on SVM has a strong applicability, and it has a considerable accuracy in the long interval recognition.

However, the modelling and classification techniques used for videos are not independent, but closely related to each other. Most existing research methods do not consider the topological structure of the space where the dynamic texture model is located, and directly classify it, thus weakening the effectiveness of the dynamic texture model. Therefore, appropriate methods should be adopted to study the topological structure of the space where the dynamic texture model is located, and then research its classification algorithm based on this. The features extracted from videos or images are often not in Euclidean space, so it is necessary to analyze these features in an appropriate manifold space. Therefore, the purpose of this paper is to study the spatial structure of dynamic textures based on the theory of Lie group manifolds, and to use SVM classification algorithms on Lie group manifold space to classify dynamic textures, and ultimately apply this technology to video classification problems.

Therefore, for video classification, this paper proposes an SVM algorithm which combines the dynamic texture features and the Lie group kernel function. The main idea is to use the Autoregressive Moving Average model (ARMA) to describe the dynamic texture features of video, map the matrix shape of Gaussians (SOG) descriptors to Lie group manifold space, and use the distance measurement on Lie group manifold as the kernel function to get the SVM classifier.

Materials and method. Video dynamic texture features and Lie group space. Video is a continuous image sequence, which has time and space related repetitive features, while dynamic texture just describes this attribute of video in motion and appearance, which can be modelled by an Autoregressive Moving Average model (ARMA), namely:

$$(1) \quad \begin{cases} \mathbf{x}(t+1) = A\mathbf{x}(t) + \boldsymbol{\nu}(t)\boldsymbol{\nu}(t) \sim N(\mathbf{0}, \mathbf{Q}) \\ \mathbf{y}(t) = C\mathbf{x}(t) + \boldsymbol{\omega}(t) \end{cases} .$$

Here, $\mathbf{x}(t)$ is state sequence, $\mathbf{y}(t)$ is observation sequence, A is transition matrix, C is observation matrix, $\boldsymbol{\nu}(t)$ is state disturbance, N is Gaussian distribution with

multivariable, $\omega(t)$ is observation disturbance, which can be set as the mean value of video sample image sequence.

After filtering the original video, the corresponding features are extracted for each frame, and the observation sequence of video clip is $Y_1^N = [\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(N)] \in R^{m \times N}$. Using the maximum a posterior probability criterion, the parameters of (1) are estimated to be [8]:

$$(2) \quad \begin{aligned} \hat{C} &= U, \\ \hat{X}_1^N &= \Gamma V^T, \\ \hat{A} &= \hat{X}_2^N (\hat{X}_1^{N-1})^\dagger, \\ \hat{Q} &= \frac{1}{N-1} \sum_{i=1}^{N-1} \hat{\psi}(t) \hat{\psi}(t-1)^T. \end{aligned}$$

Here, U, Γ, V are singular value decomposition of $Y_1^N = U \Gamma V^T$, $\hat{X}_1^N = [\hat{\mathbf{x}}(1), \hat{\mathbf{x}}(2), \dots, \hat{\mathbf{x}}(N)]$ is estimation matrix of state variable, $(\hat{X}_1^{N-1})^\dagger$ is Moore–Penrose generalized inverse of $\hat{X}_1^{N-1}$, $\hat{\Psi}(t) = \hat{\mathbf{x}}(t+1) - \hat{A}\hat{\mathbf{x}}(t)$.

The parameters $\hat{A}, \hat{C}, \hat{Q}$ of dynamic texture model represent the intrinsic attributes of the video. The recognition and classification of the video can be realized by the recognition and classification of dynamic texture parameters. Dynamic texture has some properties of Gaussian characteristics. The state $\mathbf{x}(t)$ of a system at any time is subject to Gaussian distribution with multivariate. Its mean value $\boldsymbol{\mu}(t)$ and covariance matrix $S(t)$ meet:

$$(3) \quad \begin{cases} \boldsymbol{\mu}(t+1) = \hat{A}\boldsymbol{\mu}(t) \\ S(t+1) = \hat{A}S(t)\hat{A}^T + \hat{Q} \end{cases}.$$

According to the mapping method from Gauss type feature to Lie group, the shape of Gaussians (SOG) matrix of state $\mathbf{x}(t)$ is:

$$(4) \quad G(t) = \begin{bmatrix} P(t) & \boldsymbol{\mu}(t) \\ 0 & 1 \end{bmatrix}, \quad P(t)P(t)^T = S(t).$$

The SOG descriptors correspond to each frame of dynamic texture, and the SOG descriptors sequence corresponds to the video clips. Therefore, the average value of SOG descriptors sequence can be defined as the feature matrix of dynamic texture of the video clip:

$$(5) \quad \bar{G} = \begin{bmatrix} \sum_{t=1}^N P(t)/N & \sum_{t=1}^N \boldsymbol{\mu}(t)/N \\ 0 & 1 \end{bmatrix}.$$

The $\bar{G}$ does not belong to Euclidean space, but belongs to Lie group manifold M . For any point $\bar{G}_i \in M$, there can be a tangent space $\eta_{\bar{G}_i}$. Through exponential

mapping, the tangent vector $\tau_i \in S_{\bar{G}_i}$ can be mapped to the geodesic starting from the point on $\bar{G}_i$ on M , that is:

$$(6) \quad \bar{G}'_i = \exp_{\bar{G}_i}(\tau_i) = \bar{G}_i^{1/2} \exp(\bar{G}_i^{-1/2} \tau_i \bar{G}_i^{-1/2}) \bar{G}_i^{1/2}.$$

There is also logarithmic mapping:

$$(7) \quad \tau_i = \log_{\bar{G}_i}(\bar{G}'_i) = \bar{G}_i^{1/2} \log(\bar{G}_i^{-1/2} \bar{G}'_i \bar{G}_i^{-1/2}) \bar{G}_i^{1/2}.$$

Therefore, the distance between two points $\bar{G}_i, \bar{G}_j$ on Lie group manifold M is:

$$(8) \quad d(\bar{G}_i, \bar{G}_j) = \|\tau_i, \tau_j\|_2.$$

Here, $\|\cdot\|_2$ is 2-norm.

The key of the distance calculation for video dynamic texture features on Lie group manifold is to find the geodesic equation. In fact, by using the tangent space mapping of unit element matrix, we can further obtain that (8) is equivalent to (9):

$$(9) \quad d(\bar{G}_i, \bar{G}_j) = \|\ln(\bar{G}_i^{-1} \bar{G}_j)\|_F.$$

Here, F is Frobenius norm of matrix.

SVM algorithm based on Lie group kernel function. By extracting the parameters of dynamic texture model, each video is mapped to a point on Lie group manifold space, and the video classification can be transformed into the operation of average SOG descriptor. SVM algorithm has good generalization ability in solving nonlinear and high-dimensional small sample patterns recognition. By constructing an optimal hyperplane, SVM can not only correctly separate the two types of patterns, but also maximize the separation between the two types. Because the average SOG matrix belongs to Lie group space, the distance calculation on Lie group manifold can be used to measure the dynamic texture more accurately. Taking it as the kernel function in SVM, the inner product operation of high-dimensional space can be transformed into the kernel function calculation of low-dimensional input space to avoid the problem of dimension disaster.

There is a set of Lie group test samples $\bar{G}_i$ ($i = 1, \dots, l$), $\bar{G}_i \in M$, $Z_i \in \{-1, +1\}$ is the corresponding label, then the Lie-SVM algorithm for two classification is as follows:

(1) The Lie group kernel function is selected to calculate the scalar product. The Lie group kernel function is:

$$(10) \quad K(\bar{G}_i, \bar{G}_j) = \exp[\lambda(\|\ln(\bar{G}_i^{-1} \bar{G}_j)\|_F + \|\ln(\bar{G}_j^{-1} \bar{G}_i)\|_F) + h].$$

Here, $0 < \lambda < 1$, h are scaling factor and offset factor, which are defined by user.

(2) For Lie group data structure, the following dual programming is solved:

$$(11) \quad \begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l Z_i Z_j \alpha_i \alpha_j K(\bar{G}_i, \bar{G}_j) - \sum_{j=1}^l \alpha_j \\ \text{s.t.} \quad & \sum_{i=1}^l Z_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l. \end{aligned}$$

Here, α_i is Lagrange multiplier, $C > 0$ is regular parameter.

(3) Discriminating the average SOG matrix $\bar{G}$ sample by classification function:

$$(12) \quad \begin{aligned} Z &= \text{sgn} \left(\sum_{i=1}^l \alpha_i^* Z_i K(\bar{G}_i, \bar{G}) + b^* \right), \\ b^* &= Z_i - k^* \bar{G}_i, \quad k^* = \sum_{i=1}^l \alpha_i^* Z_i \bar{G}_i. \end{aligned}$$

Considering the multi-classification patterns, based on the two classification Lie-SVM algorithm, the one against one structure is selected to design the multi-classifier (i, j represents the number of classification categories). The $(\text{Lie-SVM})_{ij}$ indicates that the classifier can distinguish the i -class and the j -class. After training each $(\text{Lie-SVM})_{ij}$, which is combined by parallel; the average SOG matrix into each classifier, and the voting rule is used to determine the final class for the output, that is, the category tag with the most votes is the class of sample, and the multi-classifier form is:

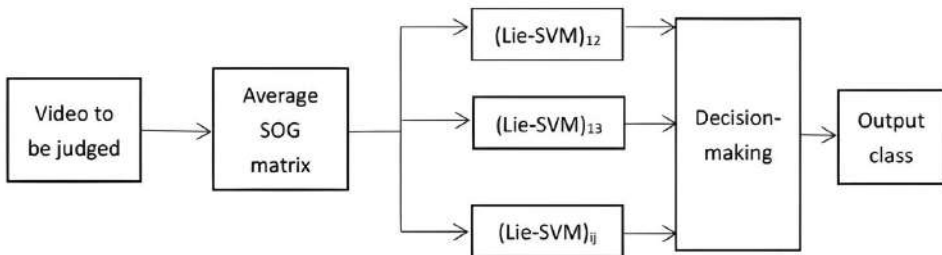
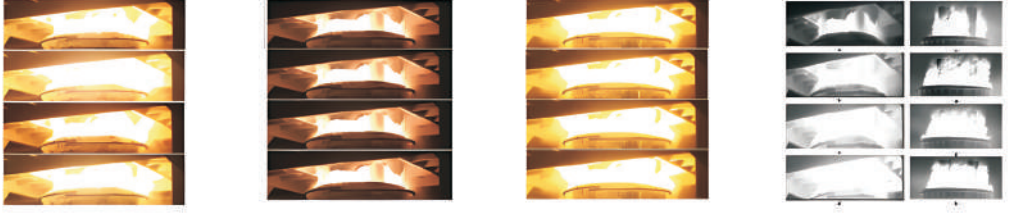


Fig. 1. Multi-classifier form

Simulation results. The source of the video data (dataset) is private. There are 140 flame video clips of electrical smelting furnace in a factory. The resolution of each video clip is 271×1051 pixels, the frames per second are 10, and each video clip has 50 frames. After the gray-scale transformation, the pixels of the video clips are all scaled to 67×262 , and the mean value is removed. There are 40 semi-fused samples, 50 feeding samples, and 50 smelting samples in 140 flame video clips. The corresponding video image sequence is shown in Fig. 2.



(1) Semi-fused condition (2) Feeding condition (3) Smelting condition (4) Reconstruction

Fig. 2. Video image sequence and reconstruction

According to [10], using the tourist behaviour graph theory, the texture features sequence of flame video can be taken as:

$$\begin{aligned}
 \mathbf{y}(t) &= \{TP_{XY}, TP_{XT}, TP_{YT}\}, \\
 TP_{XY} &= \{TP_{p,rule}^{XY}(p+1), TP_{p,rule}^{XY}(p+2), \dots, TP_{p,rule}^{XY}(p+m)\}, \\
 TP_{XT} &= \{TP_{p,rule}^{XT}(p+1), TP_{p,rule}^{XT}(p+2), \dots, TP_{p,rule}^{XT}(p+m)\}, \\
 TP_{YT} &= \{TP_{p,rule}^{YT}(p+1), TP_{p,rule}^{YT}(p+2), \dots, TP_{p,rule}^{YT}(p+m)\}.
 \end{aligned}
 \tag{13}$$

Here, $TP_{p,rule}^{XY}$, $TP_{p,rule}^{XT}$, $TP_{p,rule}^{YT}$ are texture distribution functions of flame video in three orthogonal planes; XY , XT , YT , p , m , $rule$ are set parameters.

Taking the state sequence dimension as 10, according to ARMA, the parameters estimate in texture features are $\hat{A}$, $\hat{C}$, $\hat{Q}$, $\hat{X}_1^{50}$. The (a), (b), (c) and (d) in Fig. 2(4) are the input images of video monitoring, and the (e), (f), (g) and (h) are the flame video sequences recovered by the estimated parameters of model.

It can be seen from Fig. 2(4) that the reconstructed visual features of virtual flame are basically consistent with the flame in the scene video. Furthermore, the SOG descriptors sequence $\{G(1), G(2), \dots, G(50)\}$ and the average SOG matrix $\bar{G}$ are obtained.

For each working condition $\bar{G}$ of flame video, 30-video clips are randomly selected as training samples and the rest as test sets. According to the Lie-SVM algorithm, the parameter $\lambda = 2$ and the regular factor $C = 1$, each classifier is trained by 30-positive samples and 30-negative samples, and then the classification results of the test sets are counted. The confusion matrix of (Semi-fused condition, Semi-fused condition) is 32, (Semi-fused condition, Feeding condition) is 5, (Semi-fused condition, Smelting condition) is 3, (Feeding condition, Semi-fused condition) is 6, (Feeding condition, Feeding condition) is 42, (Feeding condition, Smelting condition) is 2, (Smelting condition, Semi-fused condition) is 4, (Smelting condition, Feeding condition) is 3, (Smelting condition, Smelting condition) is 43. The recognition rates are shown in Table 1.

It can be seen from Table 1 that although there are some misjudgments in flame videos under various working conditions, the recognition effect of smelting

T a b l e 1

Classification results of flame video under various working conditions

Target type	Lie-SVM		Martin SVM from [11]	
	Time/s	Accuracy	Average time	Average accuracy
Semi-fused condition	253.43	80%	260.44	72.30%
Feeding condition	254.52	84%		
Smelting condition	255.56	86%		

flame is the best. The video dynamic texture features classification accuracy based on Lie-SVM algorithm is improved by more than 11% due to the more accurate distance calculation, and the distribution of matrix data samples does not need to be known in advance, which is completely unsupervised classification. The recognition rate shows that the dynamic texture features is an effective feature level fusion method, and the Lie-SVM algorithm can effectively identify different categories of video dynamic texture.

In the process of video dynamic texture features extraction, the selection of state sequence dimension d has a certain influence on the final classification accuracy. Each time, 30-video clips are randomly selected as samples, and the rest are used as test sets. On the basis of four experiments, the average recognition rate of each working condition of video is counted. The relationship between the final classification recognition rate and state dimension is shown in Fig. 3(a).

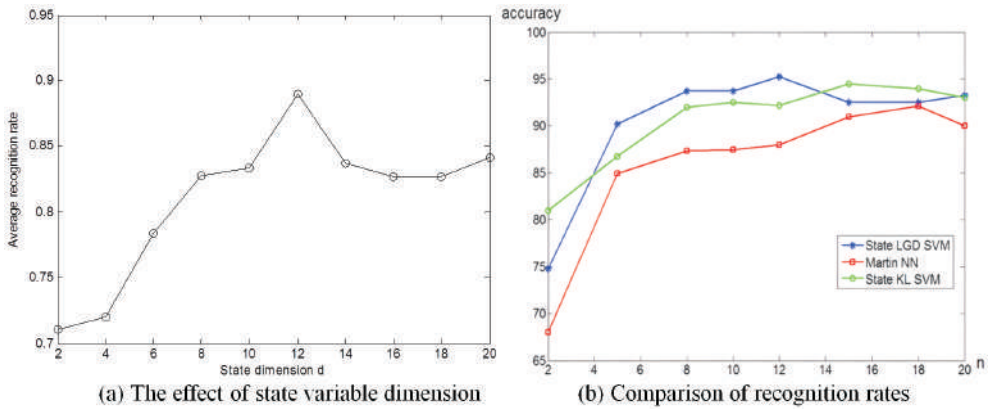


Fig. 3. The results

As can be seen in Fig. 3(a), the ARMA requires the state sequence dimension must be less than or equal to the observation sequence dimension, but this does not mean that the larger the dimension of the state sequence, the better the result. When the state sequence dimension increases to a certain extent, the classification recognition rate begins to decline, which is mainly due to the fact that the state variable corresponding to the smaller eigenvalue of singular value decomposition

is mapped onto the Lie group manifold space, and its influence on the distance between the metric matrices is limited. The best recognition in this experiment is at the position with 12 dimension.

Results and discussion. The main work summary is as follows: We find that Lie-SVM differs from or improves upon standard SVM or other classifiers in handling video texture data. The main reason is that the distance measurement methods for dynamic texture models are different. The Lie-SVM algorithm analyzes the topological space of dynamic texture models and uses the measurement method on Lie group manifolds to measure dynamic textures, while the standard SVM algorithm uses Euclidean distance. The experimental results show that: 1) the dynamic texture feature vector of flame video based on traveller's behaviour graph theory does not depend on the detection of suspected flame region, and the computational complexity is low; 2) the average SOG matrix is mapped to Lie group manifold space, and the measurement method of dynamic texture features is given, which is an indirect measurement and can effectively become a kernel function of SVM; 3) some parameters of Lie-SVM algorithms, such as scaling factor, offset factor and regularization factor, can be adjusted to adapt to the different target and environment in video recognition. Table 1 shows that the improvement of classification accuracy based on Lie-SVM algorithm by more than 11% is in average on the basis of three different conditions, and through the influence of the dimension of state sequence on the average recognition rate, it can be concluded that the optimal condition of Lie-SVM algorithm is the singular value decomposition of observation sequence with 12 dimension.

Comparison with other works. This section compares the proposed work with other studies in the field of identification of working conditions of electric melting magnesium furnace through flame video. The comparison highlights the recognition rates of three methods: Martin NN, State KL SVM (Kullback-Leibler Support Vector Machine), which means support vector machine using KL entropy as kernel or metric for dynamic texture classification/recognition in Gaussian modelling, and State LGD SVM (Lie Group based Dynamic texture Support Vector Machine), which means using Lie groups to represent dynamic textures and constructing SVM for classification. Martin measures the spatial angle of dynamic texture parameters. Since Martin distance cannot form an effective support vector machine classification kernel, the Nearest Neighbours (NN) classification method is used. KL entropy measures dynamic textures directly as Gaussian models, while LGD measures them based on Lie group manifolds, KL and LGD all use the classification method of SVM.

In Fig. 3(b), it can be seen that the results obtained based on the Lie group metric are far superior to those based on the Martin metric, and the performance is slightly better in most state space dimensions than the KL based metric. The method based on the State LGD SVM achieved a recognition rate of 95.28% at the peak, while the State KL SVM achieved a recognition rate of 94.5% at the

peak. Therefore, whether comparing the average recognition rate or the highest recognition rate, the State LGD SVM is superior to the State KL SVM and Martin NN.

Conclusions. This paper mainly studies and implements a video classification algorithm based on Lie group theory and dynamic texture. Due to the existing methods of modelling and classifying videos using dynamic textures, which ignore the spatial structural characteristics of dynamic textures, the distance measurement of dynamic texture models is inaccurate. This paper uses the theory of Lie group manifolds to analyze the structure of dynamic textures and proposes a method of mapping dynamic textures onto Lie group manifolds. Finally, this Lie group metric is constructed as a kernel function and applied to SVM learning. Experiments are conducted on a dataset of flame classification to verify the accuracy and efficiency of the Lie-SVM algorithm.

The framework of video classification algorithm based on dynamic texture and Lie group manifold mainly includes extracting features, building dynamic texture model, mapping dynamic texture to Lie group manifold, distance measurement on Lie group and recognition with SVM.

In the future, research initiatives will focus on the SVM algorithm on Lie group manifold and can be extended to video classification with multi-features fusion. Meanwhile, we can combine the kernel formed by measuring dynamic textures based on Lie group manifolds and Stiefel manifolds.

REFERENCES

- [1] BREZEALE D., D. J. COOK (2008) Automatic video classification: A survey of the literature, *IEEE Transactions on Systems, Man, and Cybernetics*, **38**(3), 416–430.
- [2] RIPAN C. (2018) Video classification method based on clustering, Huaibei: Huaibei Normal University, (in Chinese).
- [3] IBRAHIM Z. A. A., M. SAAB, I. SBEITY (2019) VideoToVecs: a new video representation based on deep learning techniques for video classification and clustering, *SN Appl. Sci.*, **1**, 560, <https://doi.org/10.1007/s42452-019-0573-6>.
- [4] HAO Y. S., Y. X. YIN, J. H. LAN (2013) Road vehicle classification based on extreme learning machine, 3rd International Conference on Electric and Electronics, Hong Kong, <https://doi.org/10.2991/eeic-13.2013.101>.
- [5] BOSER B. E., I. M. GUYON, V. N. VAPNIK (1992) A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory*, 144–152, <https://doi.org/10.1145/130385.130401>.
- [6] WATCHARAPINCHAI N., S. ARAMVITH, S. SIDDHICHAI, S. MARUKATAT (2007) A discriminant approach to sports video classification, *Proceedings of Communications and Information Technologies*, 557–561, <https://doi.org/10.1109/ISCIT.2007.4392081>.
- [7] MAJI S., A. C. BERG, J. MALIK (2013) Efficient classification for additive kernel SVMs, *IEEE Trans. Pattern Anal. Mach. Intell.*, **35**(1), 66–77.

- [8] SHOUGANG C. (2013) Research on video classification technique based on Lie group and dynamic texture, Wuhan: Huazhong University of Science & Technology, (in Chinese).
- [9] YUN Z. (2011) Mean-shift based covariance matrices clustering on differentiable manifolds, Qinghuangdao: Yanshan University, (in Chinese).
- [10] YAQIN Z. (2015) Dynamic texture analysis and recognition of flame video using tourist walk graph, Computer Engineering and Design, **36**(10), 2770–2775, (in Chinese).
- [11] LEI Z., L. SHAOWEN, Z. XIUPING (2019) Conditions recognition of fused magnesia furnace based on flame dynamic texture, Control Theory & Applications, **36**(9), 1565–1572, (in Chinese).

¹*School of Electronics and Internet of Things,
Chongqing Polytechnic University of Electronic Technology,
Chongqing 401331, China
e-mail: 328886771@qq.com*

²*College of Electronic and Information Engineering, Southwest University,
Chongqing 400715, China
e-mail: wuxizhiye@163.com*

³*School of Mechanical and Electrical Engineering, Huainan Normal University,
Huainan 232038, China
e-mail: zhmlwxcstu@163.com*