

# Image Recognition and Classification based on Elastic Model and BOF Algorithm

Mingzhu Liu<sup>\*</sup>, Xue Bao, and Lu Pang

*The Higher Educational Key Laboratory for Measuring and Control Technology and Instrumentations of Heilongjiang Province, Harbin University of Science and Technology, Harbin, 150080, China*

---

## Abstract

To solve the problems of mosaic effect and block distortion in the process of image enlargement or reduction, which decrease the accuracy of image recognition and classification, a new fusion algorithm is proposed in this paper. It is a method that integrates an elastic model with the BOF algorithm. Firstly, the distortion phenomenon of static images is studied in depth. Then, the basic principle of the classical BOF algorithm is studied. At the same time, the method and steps of generating feature descriptors using the BOF algorithm are deeply analyzed, and the clustering of feature elements is realized by using the spatial pyramid method. Aiming at the problem of low classification accuracy of scaled images with block distortion, an elastic model method is proposed. It introduces elasticity parameters to scaled image processing and combines the new features obtained by the BOF algorithm to further compensate the block distortion, so as to improve the image recognition accuracy.

**Keywords:** image recognition; block distortion; BOF algorithm; elastic model

(Submitted on September 11, 2019; Revised on September 30, 2019; Accepted on October 25, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

In recent years, computer vision has become a popular research field. Computer vision involves studying how to use computers to simulate biological vision. Its main task is to visually observe the world like human beings by processing pictures or videos collected by computers, with the ability to identify, understand, and adapt to the environment. As an important part of image processing and robot vision, feature extraction and classification technology can efficiently process visual information, obtain information needed by people, and bring convenience to life and industrial production [1]. Examples of computer vision applications include autonomous vehicle visual navigation based on the classification and recognition environment of images, identification and classification of satellite remote sensing images, specific target recognition of industrial robot hand-eye systems, and identification of biological characteristics. Accurately obtaining and processing the information required by users in vast amounts of data has become an urgent problem in the field of computer vision.

There are many methods for the classification and recognition of images, such as statistical recognition methods, structure recognition methods, neural network methods, fuzzy set recognition methods, support vector machines, and template matching methods. Different methods have their corresponding advantages and disadvantages. Support vector machine is a new artificial intelligence learning algorithm, and it has low requirements of templates. If the number of templates is small, the obtained classification result is the optimal solution under the existing template information. According to the experimental requirements of this paper, the SVM method is selected to perform image classification and recognition.

## 2. Related Work

Image recognition classification is generally divided into the following steps: inputting test images, performing image preprocessing, performing image segmentation, extracting image features, performing image recognition, performing image

---

<sup>\*</sup> Corresponding author.

E-mail address: [lmz@hrbust.edu.cn](mailto:lmz@hrbust.edu.cn)

classification, and outputting the classified results [2].

At present, commonly used methods for processing images include the deep convolutional neural network (DCNN), bag of features (BOF), K-means, SVM coupling algorithm (KM-SVM), and related improved algorithms [3]. Shi, Zhu [4], and others used the BOF model and spatial pyramid matching to solve image-based weather state recognition problems and used SVM for classification. There are few studies on the impact of pre-training and fine-tuning on the performance of convolutional neural networks in China. Duan and Zhai [5] used a more intuitive analysis of the computational process to visualize some of the hidden layer features in the convolutional network. The classification effect of deep convolution under random initialization and pre-training model initialization conditions was analyzed for the Caltech-101 dataset, and the global fine-tuning mode and local fine-tuning mode effectively optimized the image recognition process in image classification. Zhang [6] studied different colour differences in images of Fujian black tea, used digital image processing technology combined with applied machine learning, and proposed a tea image recognition method based on K-means and the SVM coupling algorithm. This paper mainly combines the BOF algorithm with the elastic model to improve the effect of image distortion caused by scale transformation on the low classification rate of the object classification recognition rate.

### 3. Method

#### 3.1. Research on BOF Algorithm

The BOF algorithm extracts feature descriptors by scale-invariant feature transform (SIFT), performs K-means clustering on the extracted feature descriptors, and finally classifies them using an SVM classifier.

The images used in this paper are from the Caltech-101 dataset. The six types of images selected in this paper are similar: lotus and sunflower, joshua\_tree and bonsai, and butterfly and dragonfly. This can reduce the chance probability of image recognition classification, so the results can be more objective. Firstly, 40 images are selected from each category as training samples. Feature descriptors are extracted by the SIFT algorithm, the word frequency is counted according to the K-means clustering algorithm, and they are finally classified by the LIBSVM software package. 20 images are selected from each of the remaining images for testing, and each test image is operated by two methods: one directly zooms in and out, while the other uses the elastic model to zoom in and out to form a comparison test library. Then, the images are identified and classified by the BOF algorithm, and the experimental results are obtained.

##### 3.1.1. SIFT Algorithm

The SIFT algorithm is a computer vision algorithm used to detect and describe local features in an image. It looks for extreme points in a spatial scale and extracts their position, scale, and rotation invariants. This algorithm was developed by David Lowe [7], published in 1999, and completed in 2004.

Lowe [7] decomposed several steps: the SIFT algorithm into scale space extremum detection, the key point location and direction determination, and the key point description.

##### 3.1.1.1. Scale Space Extreme Point Detection

The SIFT algorithm searches for key points in different scale spaces, and the acquisition of scale space needs to be implemented using Gaussian blur. The Gaussian convolution kernel is the only transform linear kernel to realize scale transformation. The scale space is represented by the difference of Gaussian pyramid, and the purpose of scale space theory is to simulate the multi-scale features of image data. Then, the scale space  $L(x, y, \sigma)$  of an image is defined as the convolution of a Gaussian function  $G(x, y, \sigma)$  with a variation scale, and the original image  $I(x, y)$  is shown in Equation (1).

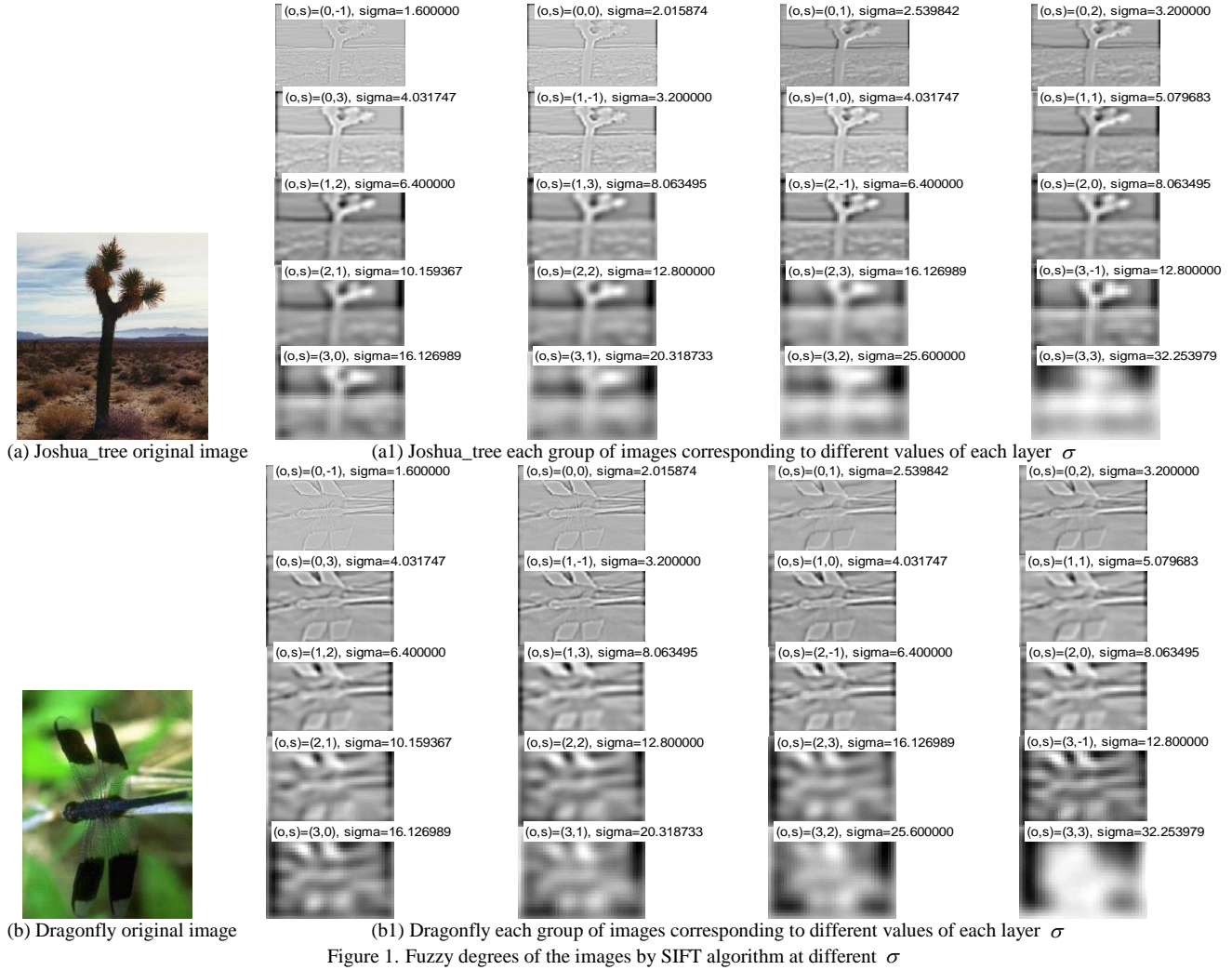
$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

Where  $*$  represents the convolution operation and  $G(x, y, \sigma)$  is represented by Equation (2).

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-m/2)^2 + (y-n/2)^2}{2\sigma^2}} \quad (2)$$

$m, n$  indicates that the dimension of the Gaussian template is determined by  $(6\sigma + 1) * (6\sigma + 1)$ , the size of the scale

space factor  $\sigma$  determines the smoothness of the image, the large scale corresponds to the outline features of the image, and the small scale corresponds to the detailed features of the image. When the value of  $\sigma$  is large, it corresponds to coarse scale (low resolution); when the value of  $\sigma$  is small, it corresponds to fine scale (high resolution). The degree of blur of each scale image in the scale space gradually becomes larger, which can simulate the formation process of the target on the retina when the target is from near to far; the larger the scale, the more blurred the image. In [7], the initial scale was set to 1.6. The initial scale adopted in this paper is also  $\sigma = 1.600000$  ( $\sigma$  is the sigma in Figure 1). Select several types of images in the Caltech-101 library used in this article (the name of the folder contained in the image is used as the image name). Figure 1 shows the ambiguity effect of the scenes of joshua\_tree and dragonfly  $\sigma$  ( $\sigma \in [1.600000, 32.2539979]$ ) on different scale factors.



In order to find the extreme points of the scale space, each check point is compared with all its neighbors to see if it is larger or smaller than the adjacent points of its image domain and scale domain. As shown in Figure 2, the middle detection point is compared with eight adjacent points of the same scale and  $9 \times 2$  points corresponding to the upper and lower adjacent scales to ensure 26 points in both the scale space and the 2D image space to the extreme point [8]. If a point is the largest or smallest value in the DOG scale space layer and the 26 fields of the upper and lower layers, the point is considered to be a feature point of the image at the scale. For the image in Figure 1, the extreme points extracted at  $\sigma = 1.600000$  are shown in Figure 3.

### 3.1.1.2. Feature Point Position

Since the DOG value is sensitive to noise, it is necessary to screen the previously selected local extreme points to remove the unstable and falsely detected points. For the extreme point distribution map in Figure 3, the accurate key point

distribution corresponding to each image can be obtained, as shown in Figure 4. The direction of the stable key point is to calculate the gradient direction of all points in the neighborhood centered on the key point. The gradient direction is  $0 \sim 360^\circ$ , normalized to 36 directions, and  $10^\circ$  is a range. The gradient histogram is generated, and the main peak is the key point, that is, a feature point. The gradient magnitude is obtained by Equation (3), and the gradient direction is obtained by Equation (4).

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (3)$$

$$\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (4)$$

Where  $m(x, y)$  is the gradient magnitude and  $\theta(x, y)$  is the gradient direction.

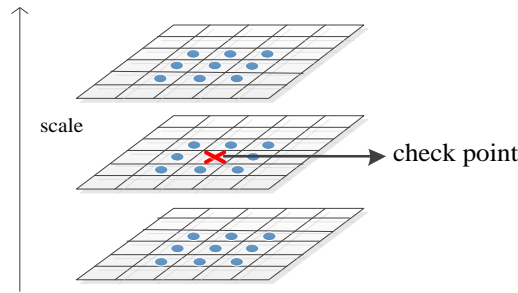


Figure 2. Schematic diagram of searching for the DOG extreme point

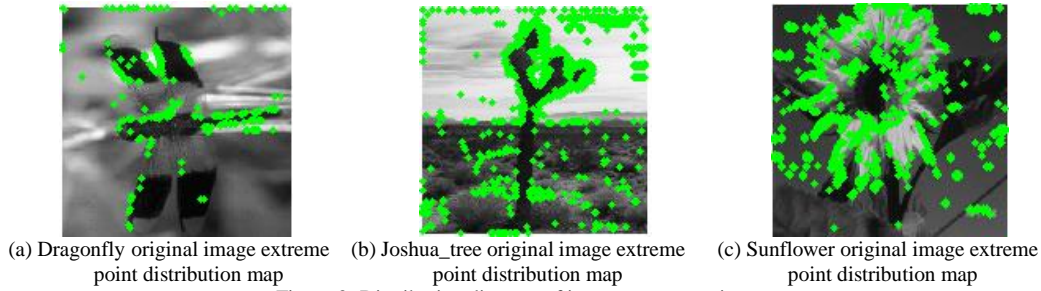


Figure 3. Distribution diagram of image extreme points

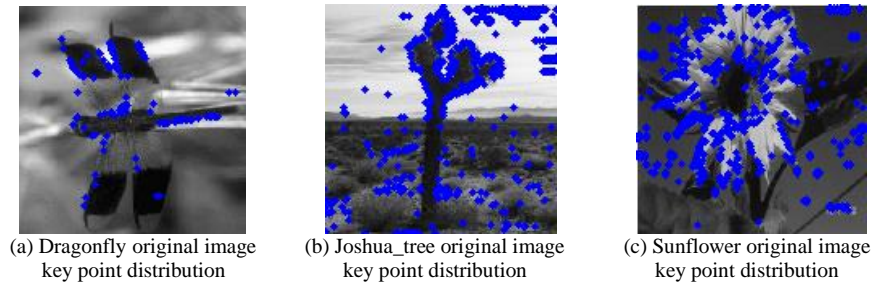


Figure 4. Distribution diagram of image key points

### 3.1.1.3. Generating Feature Descriptors

Taking  $4 \times 4$  blocks around the feature points, each block takes eight gradient directions to form  $(4 \times 4) \times 8 = 128$  dimensional feature vectors. It generates corresponding feature descriptors and obtains a large number of features to represent the keywords of the image [9].

The feature descriptor generated by each image is used as the key of the image for the next clustering, as shown in Figure 5.

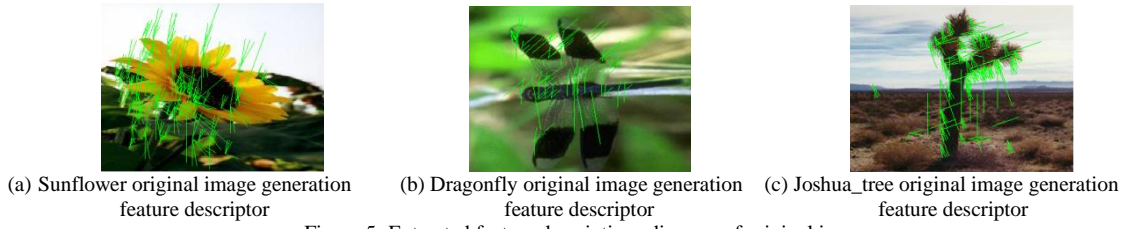


Figure 5. Extracted feature descriptions diagram of original image

### 3.1.2. K-Means Clustering Method and Visual Dictionary Acquisition

The K-means clustering algorithm [10] is a typical distance-based similarity evaluation index. The closer the distance between two targets, the greater the similarity. The algorithm considers that the cluster is composed of targets close to each other [11]. Therefore, the compact and independent cluster is the ultimate goal.

Let  $n$  training samples be  $\{x_1, x_2, \dots, x_n\}$ , ( $x_i \in R^n, i \leq n$ ) to cluster the samples into  $k$  clusters. Firstly, randomly select samples as the initial cluster centroids, and the centroid set is  $C = \{c_1, c_2, \dots, c_k\}$ , ( $k \leq n$ ). For each sample  $x_i$  in the dataset, calculate its distance to the  $k$  cluster centers according to Equation (5) and assign it to the class corresponding to the cluster center with the smallest distance.

$$\min \sum_{i=1}^k \sum_{x \in C_i} \text{dist}(c_i, x)^2 \quad (5)$$

Where  $\text{dist}(\cdot)$  is the Euclidean distance, and for each category  $c_i$ , the centroid of the class is recalculated according to Equation (6).

$$c_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (6)$$

Constantly iterate the centroid until it does not change.

In the BOF algorithm, the image is obtained by the SIFT algorithm to obtain keywords for each image. These keywords need to be clustered by K-means clustering, and the obtained cluster center is defined as a visual word. All cluster centers constitute a visual dictionary, with subsequent recognition processing.

### 3.1.3. Space Pyramid Matching

The image classification method based on BOF involves weighting the descriptor vector of each local block of the image to describe the whole image. Since there are many feature descriptors in the image, cluster quantization is required to be a "word", one image is composed of multiple words, the word frequency is formed into a histogram as the expression of the image, and finally the image is classified by the SVM classifier [12]. However, the main problem with this method is that the combination of words with low abstraction level does not fully reflect the entire image, so the spatial pyramid is introduced to convert the local conversion block into words while retaining the spatial information of the image block.

The spatial pyramid matching mainly includes the pyramid matching kernel and the matching scheme of the spatial pyramid under the BOF algorithm [13]. By dividing the image into different levels, the histogram representation based on visual words is performed, the matching points under each level are obtained according to Equation (7), and the matching degree between the feature vector levels is obtained according to Equation (8) [14].

The first few layers of the hierarchy are represented by  $\text{level}(i)$ , ( $i = 0, 1, \dots$ ). Divide the image of  $\text{level}(i)$  into  $\text{pow}(4, i)$  bins (each cluster center obtained by K-means clustering is  $\text{bin}$ ).

Suppose there are two feature sets  $X, Y$ , where each feature has a dimension  $d$ . Divide the feature space into different

scales  $0, \dots, L$ , where  $l \in L$ . Divide each dimension of the feature space by  $2^l$  bins at scale  $l$ , and then the feature space of the  $d$  dimension can divide  $D = 2^{dl}$  bins.

In  $level(i)$ , if the points  $x$  and  $y$  fall into the same bin, the points  $x$  and  $y$  match. The number of matching points in each bin is  $\min(X_i, Y_i)$ , where  $X_i$  and  $Y_i$  represent the  $i^{\text{th}}$  bin in the corresponding level.  $H_X^l$  and  $H_Y^l$  represent the histogram features of  $X$  and  $Y$  under  $level(l)$ .  $H_X^l(i)$  and  $H_Y^l(i)$  represent the number of feature points in  $level(l)$  where  $X$  and  $Y$  fall into the  $i^{\text{th}}$  bin, and then the total number of points matching under  $level(l)$  is calculated as shown in Equation (7).

$$I(H_X^l, H_Y^l) = \sum_{i=1}^D \min(H_X^l(i), H_Y^l(i)) \quad (7)$$

In the following,  $I(H_X^l, H_Y^l)$  is abbreviated as  $I^l$ .

The total number of matching  $I^l$  at each scale is counted. The matching under different scale matching should be given different weights, and the defined weight is  $\frac{1}{2^{L-l}}$ . The matching degree of the two point sets is calculated according to Equation (8).

$$\begin{aligned} k^L(X, Y) &= I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I^l - I^{l+1}) \\ &= \frac{1}{2^L} I^0 + \sum_{l=1}^L \frac{1}{2^{L-l+1}} I^l \end{aligned} \quad (8)$$

#### 3.1.4. LibSVM Classifier

In this paper, the classification of samples is processed by the traditional LibSVM classifier. SVM starts by analyzing the characteristics of the sample and predicts the attribute category of the object. The support vector can minimize the empirical error, and the geometric edge is the largest classification process [15]. LIBSVM is a simple, easy to use, fast, and efficient SVM pattern recognition and regression software package [16]. It provides several default parameters, and it runs faster [17].

#### 3.1.5. Elastic Model

The reduced image based on the integral multiple of the elastic model is equivalent to the process of slowly stretching the stretched elastic film. The pixel of each point is not only related to the current position, but also the pixels around it; the closer to the pixel, the greater the association. Assume that the horizontal and vertical directions of the image are reduced by an integer multiple of  $Radio$  times. For each row of pixels in the horizontal direction of the image, after calculating the compression using Equation (9), the corresponding pixel value of the row is

$$\begin{aligned} I(x/Radio, y) &= [f(x, y) + \sum_{i=1}^{Radio-1} f(x-i, y) \times (Radio-i)/Radio \\ &\quad + \sum_{j=1}^{Radio-1} f(x+j, y) \times (Radio-j)/Radio] / Radio \end{aligned} \quad (9)$$

The pixel point  $(x, y)$  in the original image corresponds to the  $(x/Radio, y)$  pixel point of the row after compression, the influence of  $(x, y)$  on itself is 1, and the other weights from the far and near points are  $1/Radio, 2/Radio, \dots, (Radio-1)/Radio$ .

Similarly, Equation (10) can be used to calculate the corresponding pixel points of each column after compression.

$$\begin{aligned}
I(x, y/Radius) = & [f(x, y) + \sum_{i=1}^{Radius-1} f(x, y-i) \times (Radius-i)/Radius \\
& + \sum_{j=1}^{Radius-1} f(x, y+j) \times (Radius-j)/Radius] / Radius
\end{aligned} \quad (10)$$

The image enlargement and reduction are different. When stretching the elastic film, it can be found that the colour of each point on the elastic film gradually changes. The colour of each point not only affects the colour of the surrounding area, but also the colour change itself. The change of colour can only be affected by the colour of the position before and after the elongation, and it is impossible to be affected by the colour at the farther position. Therefore, each row element of the original image is calculated according to Equation (11) [18].

$$f(i, j) = \frac{\left[ \frac{f(i, j-1)}{2} + 2 \times f(i, j) + \frac{f(i, j+1)}{2} \right]}{3} \quad (11)$$

Each column element is calculated according to Equation (12).

$$f(i, j) = \frac{\left[ \frac{f(i-1, j)}{2} + 2 \times f(i, j) + \frac{f(i+1, j)}{2} \right]}{3} \quad (12)$$

Since the "mosaic" or block effect is prone to occur during image enlargement, the interpolation in the amplification process is unreasonable, and the pixel value after the processing of Equations (13) and (14) can be preceded by the pixel and the back. The pixels are associated so that they are reasonably interpolated.

The processed image is enlarged, and assuming that both the horizontal direction and the vertical direction are enlarged by *Radio* times, the pixel values in the horizontal direction are calculated by Equation (13).

$$I(i, y) = I(i, y \setminus Radius) \times (1 - y \% Radius / Radius) + I(i, y \setminus Radius + 1) \times (y \% Radius / Radius) \quad (13)$$

Where "\" stand for "division" operation and "%" stands for "surplus" operation. Equation (13) is an interpolation of pixel points between two pixel points of the same row in the original image in the target image. The weight of the pixel point that is close to the interpolation point or already exists is large, the weight is far, and the value is small.

Description: where  $I(i, y \setminus Radius)$  is the gray value of the pixel in the original image before the insertion point,  $I(i, y \setminus Radius + 1)$  is the pixel point in the original image after the point to be inserted, and  $1 - y \% Radius / Radius$  and  $y \% Radius / Radius$  are the weights calculated from the distance from the point to be inserted.

Calculate the corresponding pixel points of the column after amplification on Equation (14) for the enlarged rows.

$$I(x, j) = I(x \setminus Radius, j) \times (1 - x \% Radius / Radius) + I(x \setminus Radius + 1, j) \times (x \% Radius / Radius) \quad (14)$$

### 3.2. Method for Image Classification using Elastic Model

In the process of enlarging and reducing the image, the image produce some "mosaic" or block distortion. The distortion of such images can be reduced by interpolating the target image. Typical applications are nearest neighbor methods and bilinear interpolation. Figure 6 shows the same enlargement and reduction of several types of images selected by the nearest neighbor method, bilinear interpolation method, and elastic model. It can be seen that the elastic model can effectively reduce the image magnification and the distortion that occurs during the process of shrinking.

The elastic model minimizes image distortion during the zooming in and out of the image [18]. Firstly, the unaltered image is retrieved and classified using the BOF algorithm. Secondly, the target test image is enlarged and reduced by a certain multiple and then retrieved and classified by the BOF algorithm. Finally, the same magnification and reduction of the image are performed by the elastic model and then retrieved and classified by the BOF algorithm. Comparing the above classification results, the method of combining the BOF algorithm and the elastic model can reduce the problem that the



target image reduces the image classification recognition rate due to size conversion. Figure 7 is an overall flow chart of the fusion algorithm.

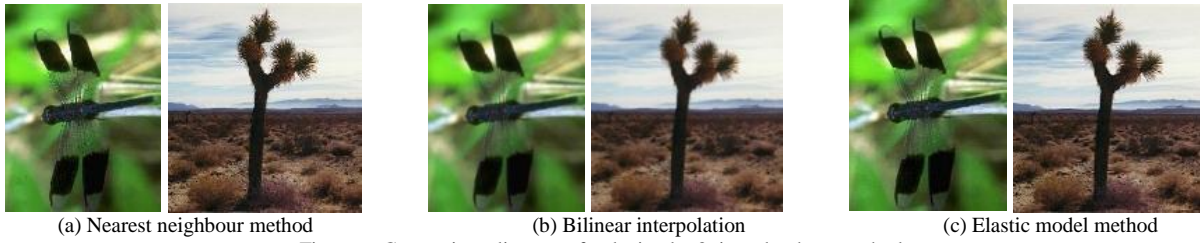


Figure 6. Comparison diagram of reducing by 8 times by three methods

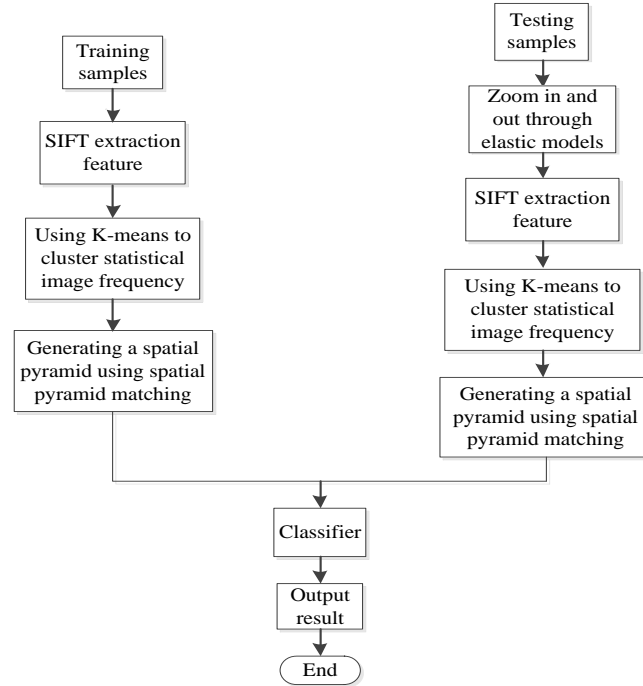


Figure 7. Flow diagram of fusion algorithm

### 3.2.1. Test for Zoom Effects

In the application process of the BOF algorithm, the classified image is directly enlarged or reduced, as shown in Table 1. The classification result can be observed, and the overall classification result can decrease. In order to reduce the influence of image scaling on image retrieval classification results, an elastic model is introduced to fuse the elastic model with the BOF algorithm. Figure 8 is a diagram of the effect of directly zooming in and out and the effect of zooming in and out of the image by the ambiguous model.

It can be seen from the picture in Figure 8 that the image causes "mosaic" and block distortion in the process of directly zooming in and out, but the enlarged and reduced pictures' "mosaic" and block after the elastic model processes the distortion is not obvious.

Table 1. Extracted number and time of feature descriptions by SIFT algorithm

Multiple of image change	Number of key points	Time (s)
Original image	91	0.085
Normally magnification 4 times	95	0.082
Normally reduced by 4 times	102	0.094
Normally reduced by 8 times	83	0.088
Elastic model magnification 4 times	65	0.081
Elastic model reduced 4 times	78	0.086
Elastic model reduced 8 times	91	0.081



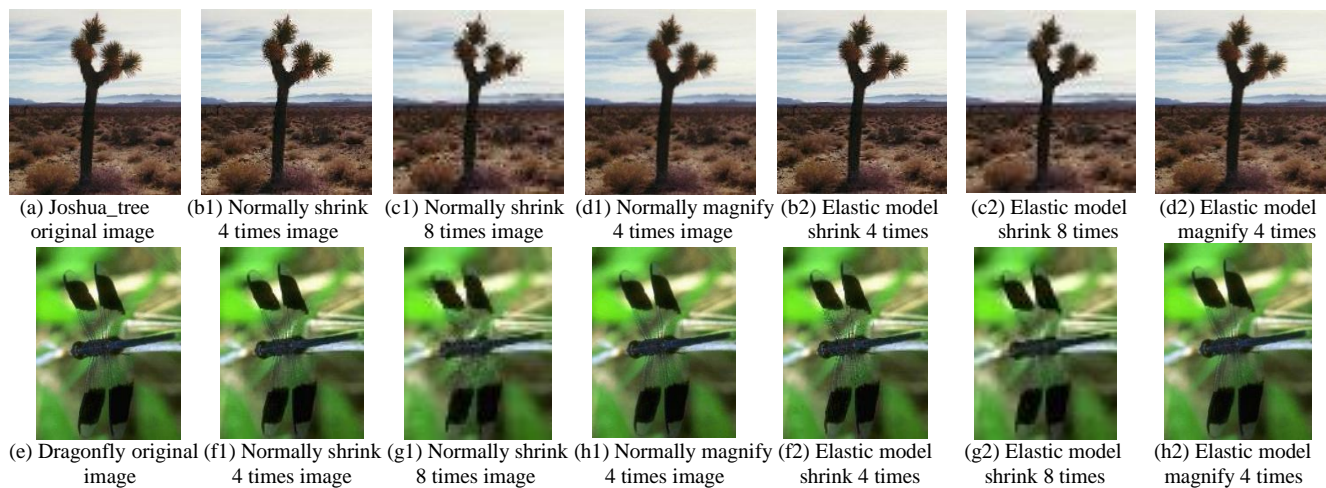


Figure 8. Comparison diagram of image scaling change

### 3.2.2. Feature Description Sub-Extraction Effect Test

The feature descriptors are extracted from all the images processed in different manners in Figure 8, the number of extracted feature descriptors is shown in Table 1, and the processed effect diagram is shown in Figure 9. It can be seen that the image after elastic model processing has fewer key points obtained by SIFT extraction feature descriptors, but according to the final recognition classification rate results, it can be seen that the final recognition and classification rate is greatly improved. It is a key point of effectiveness. It can also be confirmed that the elastic model can effectively reduce the image distortion caused by the "mosaic" and block distortion caused by enlargement or reduction.

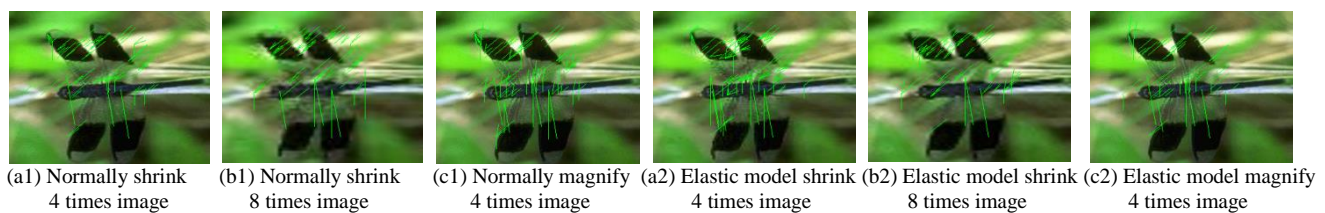


Figure 9. Effect diagram of generate feature description

## 4. Experiments

In this paper, the data in the image recognition classification library Caltech-101 database is used. Because the data volume and categories of the library are sufficient, the research used in this topic is more representative. This topic selects six types of images. Each of the papers uses 40 pictures for training, 20 pictures for testing, and a total of 240 pictures.

Figure 10 shows the results of the classification of the six types of contrast image libraries using the fusion model and the BOF algorithm and the BOF algorithm alone. It can be seen that the introduction of the elastic model can reduce image distortion during the process of zooming in and out of the image, and it can improve the recognition and classification effect of such images.

Table 2 enlarges each type of image by four times, shrinks by four times, and shrinks by eight times using the ordinary method, nearest neighbor method, bilinear interpolation method, and elastic model, and then it classifies each processed image using BOF algorithm comparative results.

From the results of the recognition classification rate in Table 2, it can be concluded that after the image is enlarged or reduced, it produces "mosaic" and block distortion, and the recognition classification rate of the image can be reduced. For example, after the image is reduced by eight times, the recognition rate of the common method is only 18.33%, and then the final accuracy rate can be improved by 30% by the method of this paper. Comparing the experimental results in Table 2, we can find that the method is relative to the traditional nearest neighbor method and double line. The sexual interpolation method can greatly reduce the influence of image enlargement and reduction on image recognition classification, and it can improve the overall recognition classification rate.

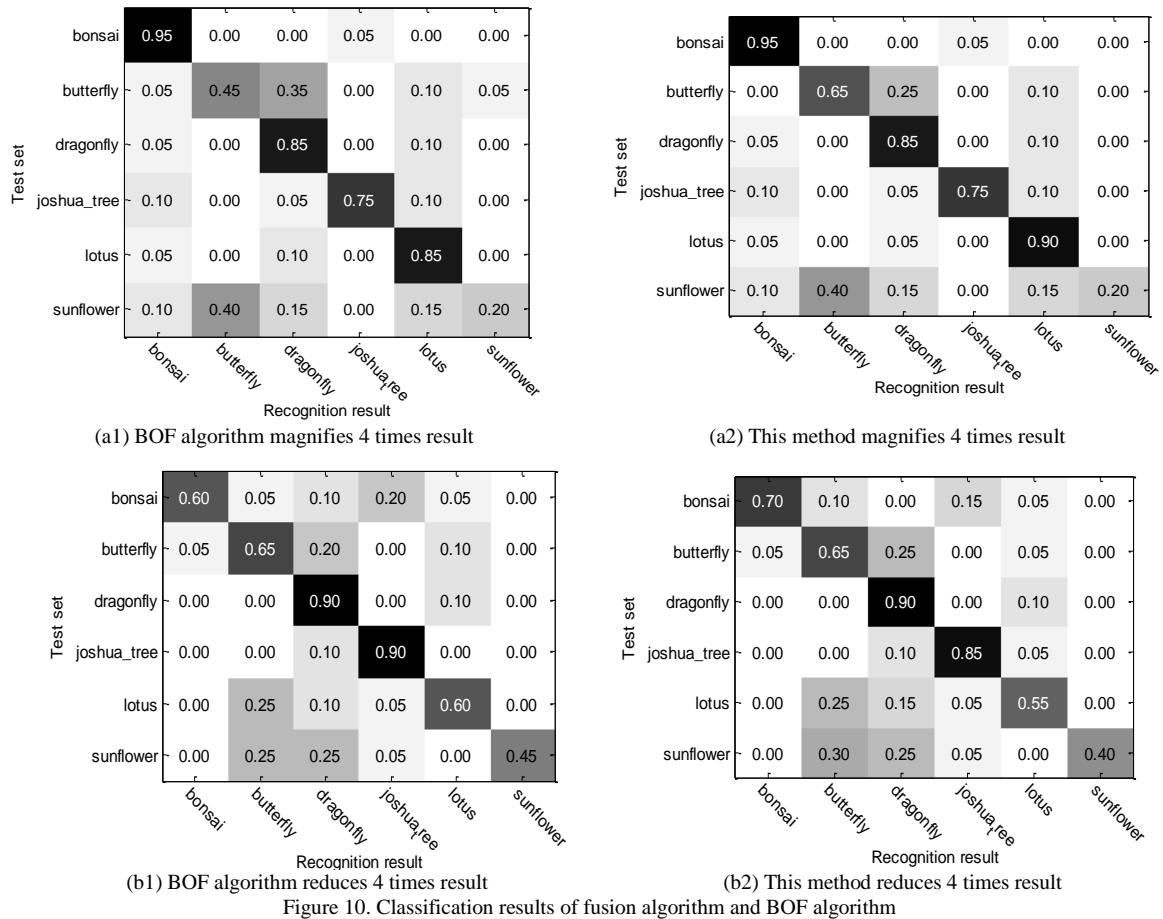


Figure 10. Classification results of fusion algorithm and BOF algorithm

Table 2. Classification results by different treatment methods

Method \ Zoom factor	Magnify 4 times accuracy (%)	Shrink 4 times accuracy (%)	Shrink 8 times accuracy (%)
Ordinary method	60.000	62.500	18.330
Nearest neighbour method	60.000	67.500	44.167
Bilinear interpolation method	69.167	65.000	46.667
Elastic model method	71.667	70.833	49.167

## 5. Conclusions

If the target classification image is directly enlarged and reduced, it can cause some mosaic phenomenon or block distortion. Therefore, using the BOF algorithm for recognition and classification of such images reduces the overall classification recognition rate. The elastic model can improve the problem of image distortion caused by size conversion. The experimental results in this paper show that the fusion elastic model and BOF algorithm improve the recognition classification rate of such distorted images.

## References

1. G. Sun, L. Liang, T. Chen, F. Xiao, and F. Lang, "Network Traffic Classification based on Transfer Learning," *Computers and Electrical Engineering*, No. 69, pp. 920-927, 2018
2. G. Sun, S. Li, Y. Cao, and F. Lang, "Cervical Cancer Diagnosis based on Random Forest," *International Journal of Performability Engineering*, Vol. 13, No. 4, pp. 446-457, 2017
3. K. S. Li, F. B. Wang, L. X. Zhang, and W. Li, "Image Recognition and Classification based on Improved BOF Algorithm," *Journal of Central South University (Science and Technology)*, Vol. 47, No. 5, pp. 1599-1605, 2016
4. J. Shi, H. Zhu, and Y. Han, "Outdoor Weather Classification," *Computer Systems and Applications*, Vol. 27, No. 4, pp. 259-263, 2018

5. J. Duan and H. M. Zhai, "Research on Deep Convolutional Neural Network for Caltech-101 Image Classification," *Computer Applications and Software*, Vol. 33, No. 12, pp. 165-168, 2016
6. J. L. Zhang, "Tea Image Recognition based on K-Means and SVM Coupling Algorithm," *Journal of Quanzhou Normal University*, Vol. 34, No. 6, pp. 48-54, 2016
7. D. G. Lowe, "Distinctive Image Features from Scale Invariant Key Points," *International Journal on Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004
8. H. B. Huang, X. L. Li, W. Y. Xiong, and X. F. Nie, "A Survey of Image Registration based on SIFT," *Software Guide*, Vol. 18, No. 1, pp. 1-4, 2019
9. X. M. Bao, X. R. Wu, and W. T. Lv, "Mark Point Positioning Method of PCB Board based on IBBS-SIFT Algorithm," *Journal of Zhejiang Sci-Tech University (Natural Sciences Edition)*, Vol. 41-42, No. 3, pp. 360-366, 2019
10. A. Likas, N. Vlassis, and J. J. Verbeek, "The Global K-Means Clustering Algorithm," *Pattern Recognition*, Vol. 36, No. 2, pp. 451-461, 2003
11. W. B. Li, H. W. Pan, P. Y. Li, X. Q. Xie, and Z. Q. Zhang, "A Medical Image Retrieval Method based on Texture Block Coding Tree," *Signal Processing: Image Communication*, Vol. 59, pp. 131-139, 2017
12. H. Sun, X. Sun, H. Wang, Y. Li, and X. Li, "Automatic Target Detection in High-Resolution Remote Sensing Images using Spatial Sparse Coding Bag-of-Words Model," *IEEE Geoscience and Remote Sensing Letters*, Vol. 9, No. 1, pp. 109-113, 2011
13. C. Zhang, J. Liu, J. Wang, Q. Tian, C. Xu, H. Lu, et al., "Image Classification using Spatial Pyramid Coding and Visual Word Reweighting," in *Proceedings of Asian Conference on Computer Vision*, Vol. 6494, Springer, Berlin, Heidelberg, 2010
14. F. Lu and J. Huang, "Beyond Bag of Latent Topics: Spatial Pyramid Matching for Scene Category Recognition," *Frontiers of Information Technology and Electronic Engineering*, Vol. 16, No. 10, pp. 817-828, 2005
15. L. Ma, X. Y. Chen, and G. C. Deng, "Evaluator-based Extraction and Selection of Thermal Infrared Harbor Image Features," *Science of Surveying and Mapping*, No. 5, pp. 87-93, 2017
16. G. Sun, T. Chen, Y. Su, and C. Li, "Internet Traffic Classification based on Incremental Support Vector Machines," *Mobile Networks and Applications*, Vol. 23, No. 4, pp. 1-8, 2018
17. Y. P. Geng and J. Y. Zhang, "Research on the Application of LibSVM in Color Image Classification," *Wireless Internet Technology*, Vol. 15, No. 16, pp. 134-136, 2018
18. M. Kang, "Research on Some Key Algorithms for the Image Processing," Xi'an University of Electronic Technology, 2009

**Mingzhu Liu** is an associate professor in the College of Measurement and Control Technology and Communication Engineering at Harbin University of Science and Technology. His research interests include intelligent information processing and machine learning.

**Xue Bao** is a master's student in the School of Measurement and Control Technology and Communication Engineering at Harbin University of Science and Technology. Her research interest is machine vision.

**Lu Pang** is a master's student in the School of Measurement and Control Technology and Communication Engineering at Harbin University of Science and Technology. Her research interest is digital image processing.