Scientific Research Publishing

Journal of Computer and Communications, 2020, 8, 26-43
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

An SLGC Model for Asian Food Image Classification

Ruoqi Wu[1], Shuai Zhao[2], Zhijian Qu[1*]

[1]Department of Computer Science and Technology, Shandong University of Technology, Zibo, China
[2]Department of Computer Science, University of York, York, UK
Email: rq.wu@foxmail.com, shuai.zhao@york.ac.uk, wanli.chang@york.ac.uk, *zhijianqu@sdut.edu.cn

How to cite this paper: Wu, R.Q., Zhao, S. and Qu, Z.J. (2020) An SLGC Model for Asian Food Image Classification. *Journal of Computer and Communications*, **8**, 26-43.
https://doi.org/10.4236/jcc.2020.84003

Received: February 2, 2020
Accepted: April 6, 2020
Published: April 9, 2020

Abstract

As a fine-grained classification problem, food image classification faces many difficulties in the specific implementation. Different countries and regions have different eating habits. In particular, Asian food images have a complicated structure, and the related classification methods are still very scarce. There is an urgent need to develop a feature extraction and fusion scheme based on the characteristics of Asian food images. To solve the above problems, we proposed an image classification model SLGC (SURF-Local and Global Color) that combines image segmentation and feature fusion. By studying the unique structure of Asian foods, the color features of the images are merged into the representation vectors in the local and global dimensions, respectively, thereby further enhancing the effect of feature extraction. The experimental results show that the SLGC model can express the intrinsic characteristics of Asian food images more comprehensively and improve classification accuracy.

Keywords

Asian Food, Image Classification, Image Segmentation, Feature Fusion, Bag of Features

1. Introduction

With the improvement of living standards, people began to pursue a more scientific and healthy diet. The food image classification is to automatically analyze the food images provided by the user through a computer and give a matching food name to further predict the user's diet and nutrient intake [1].

Since the 1990s, relevant research on food identification has appeared. SVM-based multi-core learning, multi-feature fusion and other methods have

*Corresponding author.

been applied by researchers in the field of food recognition [2]. Hongsheng He *et al.* present an automatic food classification method, DietCam, which specifically addresses the variation of food appearances [3]. Shota Sasano *et al.* propose to characterize the color and texture information by incorporating the strategy of patch-based bag of features model, which can greatly improve the accuracy of classification [4]. Since the food images mostly show the table scene, so there are inevitably hidden objects such as tableware, condiments, tablecloths, which increases the complexity of the stage. Such problems lead to a lack of clarity in the food subject, which hurts the extraction of features, and seriously affects the effect of image classification. Also, convolutional neural networks have become a very effective method in the field of computer vision [5] [6] and are increasingly being used in the area of food image classification [7]. Takumi Ege *et al.* apply Faster R-CNN to food photos of multiple dishes and use Faster R-CNN as a food detector to detect each dish in a food image, then they estimate food calories from a food photo of multiple dishes [8]. Shu Naritomi *et al.* implement Japanese food category transformation in mixed reality using both image generation and HoloLens [9]. In the use of convolutional neural networks for image classification, to achieve better accuracy, it is necessary to provide an extensive dataset during the training process. However, the current data collection and processing for Asian food images are progressing slowly, making it challenging to implement a large-scale deep learning training process. On the other hand, the image classification method based on deep learning is highly accurate, but the results are often lack of explanatory, which is not conducive to the in-depth analysis of Asian food-specific structures and feature extraction methods [10].

From the perspective of food attributes, European, American and Asian foods differ significantly in terms of structure, morphology, texture, and colour [11]. In the West, most of the food is well structured, and the cooking style is relatively monotonous. However, Asian foods have different shapes, unclear structures, and the appearance of dishes under different cooking methods varies greatly, so it is necessary to develop an image classification scheme suitable for Asian food.

We proposed an image classification model SLGC (SURF-Local and Global Color) that combines image segmentation and feature fusion. First, the GrabCut algorithm is used to segment the original image, and extract the food subject from the image. Then, we made improvements to the BoF (Bag of Features) model, to extract the SURF (Speeded Up Robust Features) feature of the image, and the colour information in the neighbourhood of the SURF feature point as the "local colour feature". The local colour feature is merged with the SURF feature and quantized. By clustering and building a feature dictionary, we can get the "image representation vector" of the original image. Finally, the "global colour feature" of the original image is extracted and merged with the image representation vector, then input the merged features into a classification model based on the SVM (Support Vector Machine) for training and classification.

2. The Theoretical Model

The SLGC model proposed in this paper includes image segmentation, image feature extraction, local and global feature fusion and classification. Figure 1 shows the framework of SLGC.

To reduce the influence of the background of the food image on feature extraction, SLGC uses the GrabCut algorithm to segment food images. GrabCut is an interactive segmentation algorithm that uses GMM (Gaussian mixture model) to estimate the colour distribution of the segmented object and background based on the specified bounding box of the segmented object [12] [13]. Equation (1) shows the energy function of the GrabCut.

$$E(\alpha,k,\theta,z)=U(\alpha,k,\theta,z)+V(\alpha,z) \tag{1}$$

α is the transparency coefficient, k is the number of GMM components, and $\theta=\{\pi k,uk,\Sigma k\}$ is the ratio, mean and covariance corresponding to each GMM component. According to the matching degree between α and pixel z, the quality of the region data item U can be measured. The smooth term V obtains the minimum value at the image boundary, thereby getting the optimal amount of the energy function E and the best segmentation scheme is determined.

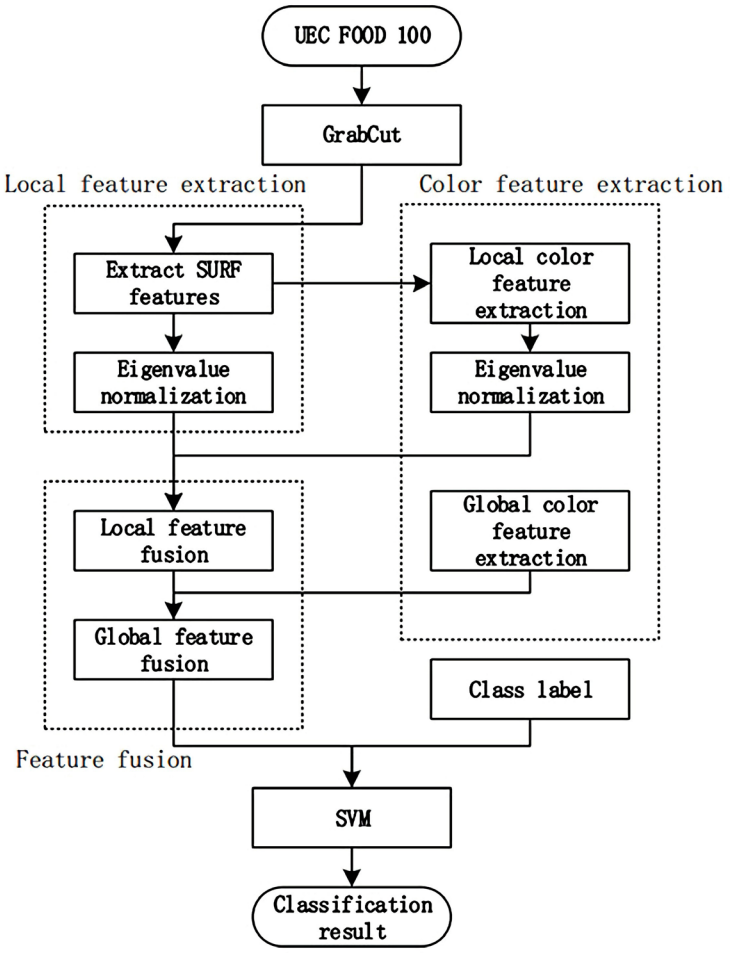

Figure 1. Framework of SLGC model.

Figure 2 shows the effect of using the GrabCut algorithm for food image segmentation. Column (a) is the original image, containing interference information such as tableware, tablecloths and other foods. Column (b) shows the effect of marking the main body of the food in the original image. Column (c) shows the effect of the image segmentation, and column (d) is the final result after cropping.

The SLGC uses SURF to extract local feature information of the image. The core of the SURF algorithm is the Hessian matrix. Using the Hessian matrix to calculate the local determinant of each pixel in the image, we can obtain the feature points of the image, as shown in Equation (2).

$$H(x,\sigma)\begin{pmatrix}\frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial xy}\\ \frac{\partial^2}{\partial xy} & \frac{\partial^2}{\partial y^2}\end{pmatrix}\cdot L(x,y,\sigma)=\begin{pmatrix}L_{xx}(x,y,\sigma) & L_{xy}(x,y,\sigma)\\ L_{xy}(x,y,\sigma) & L_{yy}(x,y,\sigma)\end{pmatrix} \tag{2}$$

$L_{xx}(x,y,\sigma)$ is the convolution of Gaussian second-order differential with the original image at point (x,y).

At the same time, to ensure the feature points that be extracted have rotation invariance, it is necessary to determine the main direction of the feature points. Counting the Harry wavelet feature in the neighbourhood centred on the feature points and radiused by six scales. Moreover, calculating the sum of the wavelet responses in the 60˚ fan window. Then the feature direction vector is obtained, as shown in Equation (3).

$$m_w=\sum_w dx+\sum_w dy,\theta_w=\arctan\left(\sum_w dx\Big/\sum_w dy\right) \tag{3}$$

m_w and θ_w represent the magnitude and direction of the feature direction

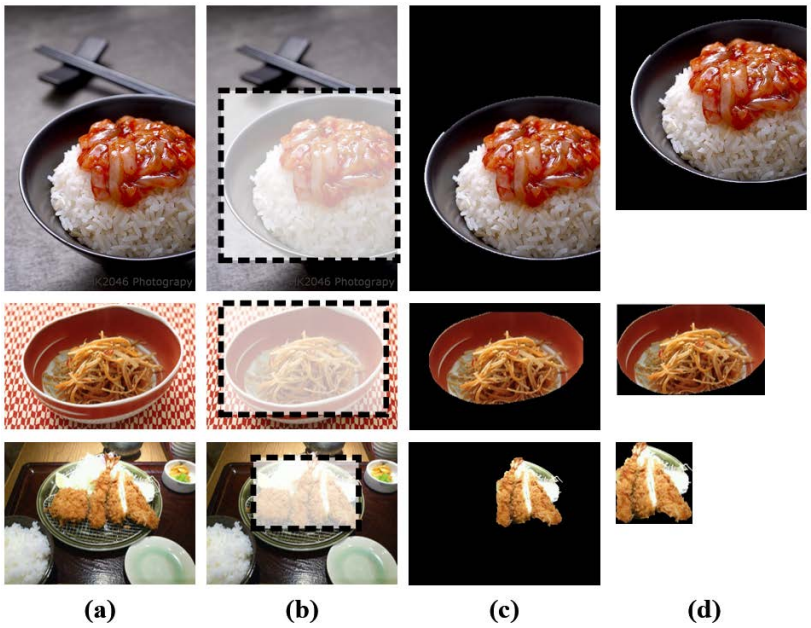

Figure 2. Process of image segmentation.

vector, respectively. Centring on the feature points, we divided the square areas of 20 scale ranges into 16 sub-blocks along the main direction. $\sum dx$, $\sum dy$, $\sum |dx|$ and $\sum |dy|$ are respectively counted to generate SURF feature descriptors, and the dimension D_s is a fixed value of 64.

Compared with SIFT (Scale-invariant feature transform), using SURF to extract the features of food images can effectively improve the speed while maintaining the appropriate number of feature points [14], which is beneficial to the subsequent real-time processing of food images.

Figure 3 shows the effect of extracting feature points on the food image using SURF. Column (a) is the original image. To make the feature point clear, set the threshold to 6000 to extract the SURF feature points and mark them in the image in column (b). The centre of the blue circle is the position of the feature point, and the different radii represent different scale information.

Figure 4 shows the change of feature vector information during local-global feature fusion.

The primary function of the feature fusion part in SLGC is to fuse the SURF feature of the food image with the local colour feature using the BoF model to enhance the representation ability of the local feature information. Then, the global colour feature of the image is added to the image representation vector to complete the global feature fusion, and the feature extraction effect is further improved. Then, the final representation vector fused by the local and global features is input into the SVM for training and classification. The following sections will focus on the process of feature fusion.

3. Feature Fusion

The critical work of the SLGC is in the feature fusion section. We made the following improvements to the BoF model. In the two steps of "feature point information quantization" and "formation of final representation vector", the local

Figure 3. SURF feature point extraction.

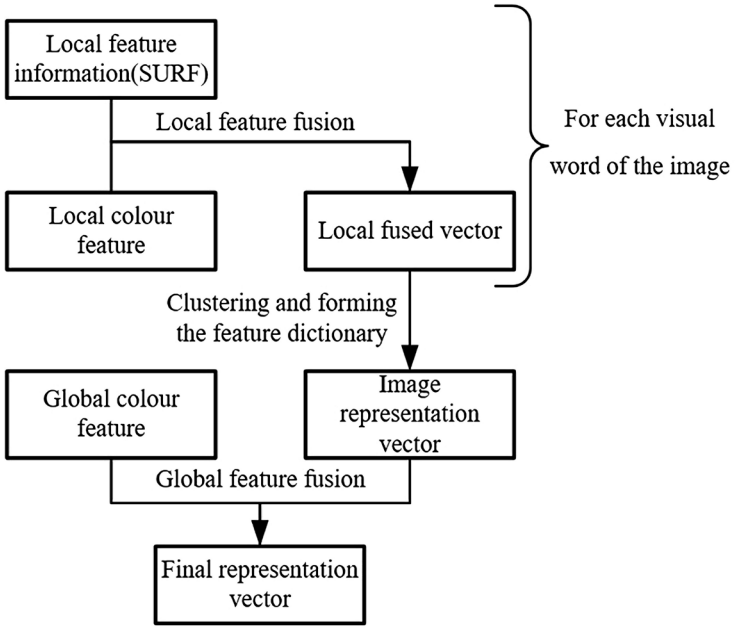

Figure 4. Change of feature information flow.

colour feature and the global colour feature of the image are respectively merged into the final representation vector. Through these two improvements, we can improve the effect of image classification effectively.

3.1. Bag of Features Model

The BoF model was proposed by Csurka and gradually applied to the field of image processing [15]. The necessary steps are as follows. Firstly, selected the training image by region, the feature points are located, and described by the feature vector respectively, as shown in **Figure 5**.

Then, the feature vector set is processed by the K-means clustering algorithm to obtain the feature dictionary. With different feature extraction methods, the number of feature points that can be located in each image is also different. If the dataset contains m images, and the number of feature points of each image is β_i, then the value of the cluster number K selected in this paper is as shown in Equation (4).

$$K = \sqrt{\sum_{i=1}^{m} \beta_i} \tag{4}$$

Finally, referring to the feature dictionary, the feature words can be extracted, and the frequency of occurrence of each word is counted to obtain the image representation vector of the original image, as shown in **Figure 6**.

3.2. Local Feature Fusion

The SURF feature point is relatively unique in the image, which can reflect the essential information of the image, so the colour information of the pixels in the neighbourhood is also critical. The SURF information is extracted and fused

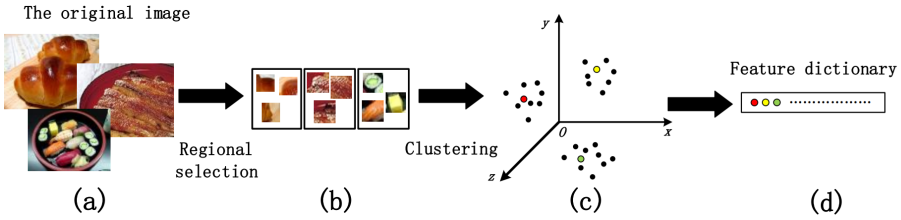

Figure 5. The generation of feature dictionary.

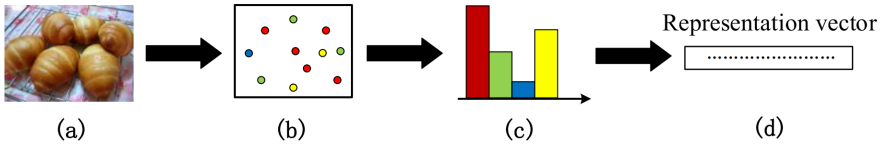

Figure 6. The generation of representation vector.

with the colour information in the neighbourhood of the feature point so that we can represent the image content more accurately and comprehensively. **Figure 7** shows an example of the feature point location and selection of neighbourhood pixel.

Point P is a certain SURF feature point, and R is the neighbourhood radius. Assume that the optimal value of R is 2 (subsequent experiments will determine the actual optimal value of R), and the RBG colour space is used to represent the colour information of pixels, which in the neighbourhood of feature points. Then, the local colour feature is formed, and we calculate its dimensions by Equation (5).

$$D_c = \left(2R^2 + 2R + 1\right) * 3 \tag{5}$$

R represents the radius of the neighbourhood and D_c represents the dimension of the local colour feature.

After obtaining the local colour feature, we improved the BoF model in the part of feature point quantization, and meantime, the local colour feature is combined with the corresponding SURF descriptor information to complete the local feature fusion. Since the feature vector and the colour information are different evaluation indicators, to eliminate the dimensional influence between the indicators, they must be normalized separately before the fusion. **Figure 8** shows the specific process of local feature fusion.

We take a visual word representing "bread" on the left as an example so that we can obtain the SURF feature vector, and the colour information in the neighbourhood of radius R. Then, we spliced the two into a local fused vector. We calculate its dimensions by Equation (6).

$$D_l = D_s + D_c \tag{6}$$

D_s represents the dimension of the SURF feature vector, which is a fixed value of 64, D_c represents the dimension of local colour feature, and D_l represents the dimension of local fused vector after feature splicing. All the local fused vectors in the image are clustered, and the feature dictionary is generated to form

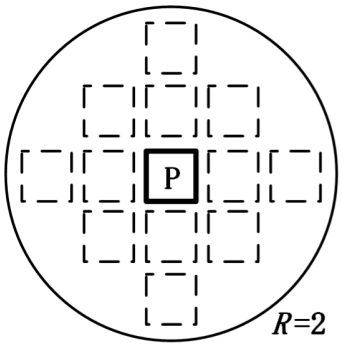

Figure 7. Selection of feature points and neighbouring pixels.

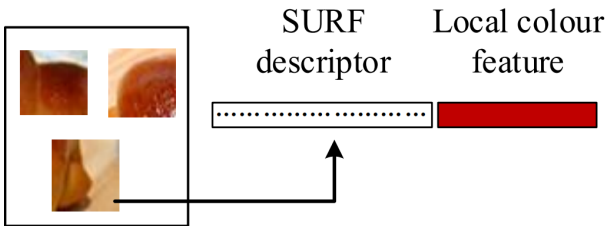

Figure 8. Local feature fusion.

the image representation vector, and the local feature fusion process is completed.

Take the image of "Udon noodles" and "pies" in the UEC FOOD 100 as an example. We marked the position of the SURF feature point in the image of column (a), and column (b) is the colour histogram corresponding to the left image. Since the two images have a large area of yellow, and the main body is similar in colour, so the degree of discrimination is low. However, by analyzing the location of the feature points, we found that most of them are located in the "embellishment area", which is the position of the red intestine and the chopped green onion. Unlike the Western habit of adorning food containers, Asian tend to embellish dishes directly with brightly coloured, fresh-tasting materials for a variety of reasons. Because the colour and shape of the embellishment area are prominent, the feature point can appear in the vicinity of it with a high probability, and the colour of the embellishment area can better reflect the type of food. By analyzing the colour information in the neighbourhood of the feature points of the two foods in **Figure 9**, it is possible to distinguish the two types of foods by the difference between red and green.

3.3. Global Feature Fusion

Before using SURF to extract the features of food images, the original images must be greyed out. The grayscaled image inevitably loses its colour features, which contains much crucial information, and they are essential for colour-rich food images. In particular, the global colour feature can represent the most widely distributed colour in the image and can play a more significant role in the classification.

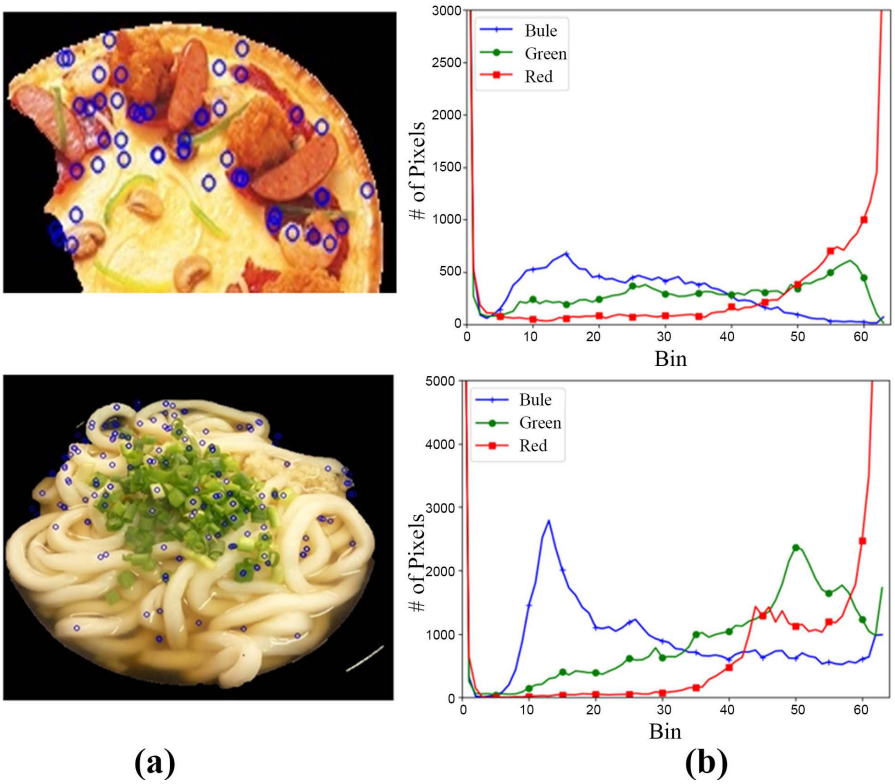

Figure 9. Food with similar main colours and their colour histogram.

This article uses the HSV colour space to represent the global colour feature of a food image. The HSV colour space is a space in which H (hue), S (saturation), and V (Value) are used as colour values to locate colours. Compared with the RGB colour space, the HSV space can intuitively express the brightness and vividness of the colour, which is closer to the natural visual perception of the food image by humans.

To avoid the vector dimension of the global colour feature being too high, we use the Equation (7) to quantify the HSV space.

$$H=\begin{cases}0 & H\in[316,20]\\ 1 & H\in[21,40]\\ 2 & H\in[41,75]\\ 3 & H\in[76,155]\\ 4 & H\in[156,190]\\ 5 & H\in[191,270]\\ 6 & H\in[271,295]\\ 7 & H\in[296,315]\end{cases}\quad S=\begin{cases}0 & S\in[0,0.2]\\ 1 & S\in[0.2,0.7]\\ 2 & S\in[0.7,1]\end{cases}\quad V=\begin{cases}0 & V\in[0,0.2]\\ 1 & V\in[0.2,0.7]\\ 2 & V\in[0.7,1]\end{cases} \tag{7}$$

Based on the above-described quantization relationship, each colour component is synthesized into a 72-dimensional colour feature vector according to Equation (8) and is used to represent the overall colour feature of the image.

$$G=9H+3S+V \tag{8}$$

To complete the global feature fusion, we improved the BoF model in the section "Formation of final representation vector", as shown in Figure 10.

The "image representation vector" is formed by clustering all the local fused vectors of the image to form a feature dictionary and complete statistics. It contains the SURF feature of the image and the colour features in the neighbourhood of the feature points. The image representation vector is spliced with the HSV global colour feature of the image to form a "final representation vector". We calculate its dimensions by Equation (9).

$$D_g = K + D_h \tag{9}$$

K is the number of clusters according to formula (4), in the meantime, it is the dimension of the image representation vector. D_h is the dimension of the global colour feature, which is a fixed value of 72, and D_g is the dimension of the final representation vector after the global feature fusion. Figure 11 shows a colour

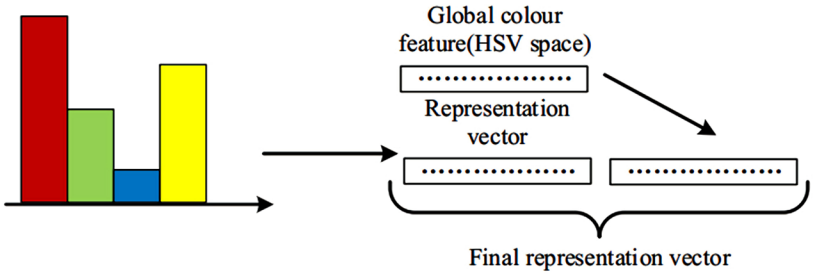

Figure 10. Global feature fusion.

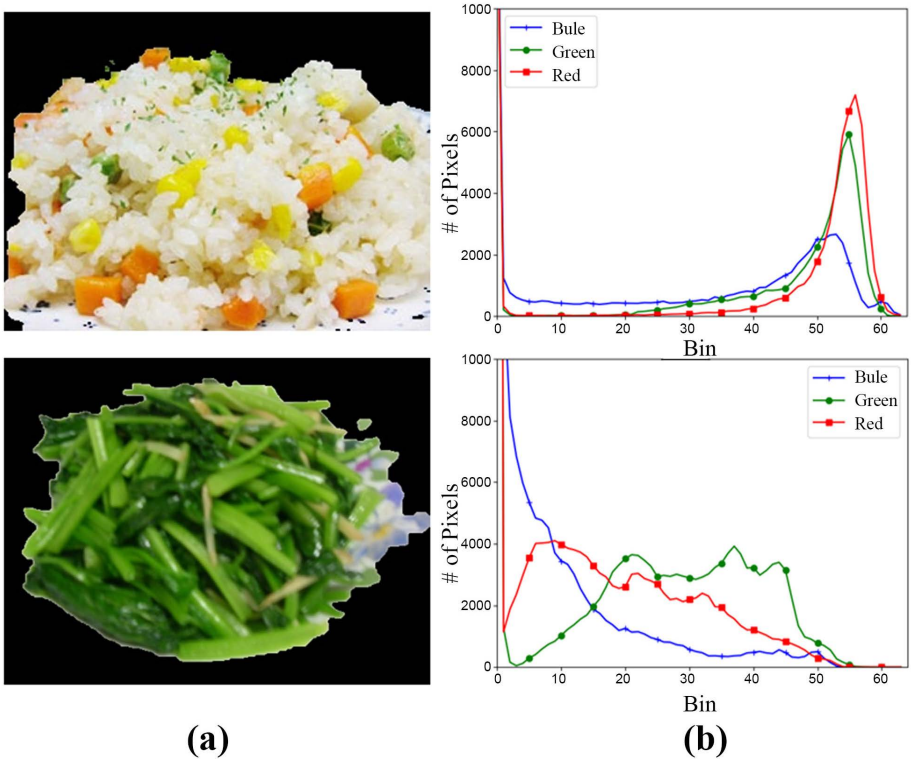

Figure 11. Foods with different main colours and their colour histograms.

histogram comparison of different foods. Through analysis, we can see that the global feature fusion is reasonable and practical.

Column (a) shows two kinds of foods, "fried rice" and "vegetables", they have different primary colour tones, which are beige and turquoise. Through the colour histogram of column (b), we find that the difference in the distribution of the two colours is significant. Therefore, global feature fusion can further improve the ability of the feature vector to express the content of food image and further improve the classification accuracy rate.

4. Experimental Results and Analysis

To verify the validity of the SLGC, we validate its key steps from two perspectives. Firstly, by comparing the classification effects before and after image segmentation to see if it contributes to the classification accuracy. Secondly, by comparing the classification accuracy under different value of neighbourhood radii, and the classification accuracy before and after the global feature fusion, we studied the influence of local feature fusion and global feature fusion on image classification separately.

4.1. Dataset

The datasets in the experiment were Caltech 101 and UEC FOOD 100. Caltech101 is an integrated image dataset with rich content and different types of images. UEC FOOD 100 is a food image dataset created by Yoshiyuki Kawano of the University of Electro-Communications, most of which are popular Japanese foods, which can fully reflect the structural characteristics of Asian food. Table 1 shows the structure of the dataset.

We use linear SVM (Radial Basis Function) as a classifier for food images and use a one-versus-rest strategy for multi-classification. Optimize parameters using the grid.py tool of Libsvm. We did not divide the training and test sets manually. Instead, K-fold cross-validation (K-CV) is used to divide and train the dataset K times and obtain K classification models. Taking the average value p_{ave} of the classification accuracy p_i of the K models as the performance of this K-CV, from this, the optimal parameters C and $gamma$ are determined, and then the image classifier is constructed using the optimal parameters to obtain the best classification accuracy P.

The experimental environment is Windows 10 operating system, Intel Core i5 CPU, 16G memory, programming environment is PyCharm 2018.2.7, Python 2.7. We modified the relevant code of OpenCV 3.3 to extract the SIFT and SURF features and used Libsvm for parameter optimization and SVM training.

Table 1. Dataset structure.

Dataset	Number of images	Number of image types	Image content
Caltech101	9145	102	Complex
UEC FOOD 100	14366	100	Food

4.2. Image Segmentation

For the problem of tablecloths, tableware and other interference information in the background of food image, the GrabCut algorithm is used to segment the image to extract the food subject. Using UEC FOOD 100 as the experimental dataset, the classification and comparison experiments before and after image segmentation were carried out for the image data of different scales. This experiment is mainly to evaluate and quantify the effect of image segmentation on classification accuracy. Figure 12 shows the experimental results.

In Figure 12, the abscissa indicates the size of different datasets (the number of image types), and the ordinate indicates the classification accuracy. By extracting the SURF and SIFT features and performing image classification experiments on the dataset before and after the segmentation, we observed that with the same experimental data, image segmentation could achieve a 4% to 6% improvement with SIFT and a 5% to 8% increase with SURF. It shows that image segmentation can effectively highlight the food subject and avoid the adverse effects of background interference on subsequent processing.

At the same time, the image segmentation effectively reduces the amount of resources used for subsequent processing. During the experiment, the size of the dataset can have a decisive impact on the efficiency of the classification model. Furthermore, we can conclude from Equation 6 that the dimension of the SURF information in the local fused vector is unchanged, and the dimension of the local fused vector depends on the value of the neighbourhood radius of the feature point. Therefore, the value of the neighbourhood radius will affect the execution efficiency of the model. As shown in Table 2, we performed image segmentation under different data sizes and the value of the neighbourhood radius. Experiment shows that the use of image segmentation can effectively reduce the time of feature extraction by about 45%.

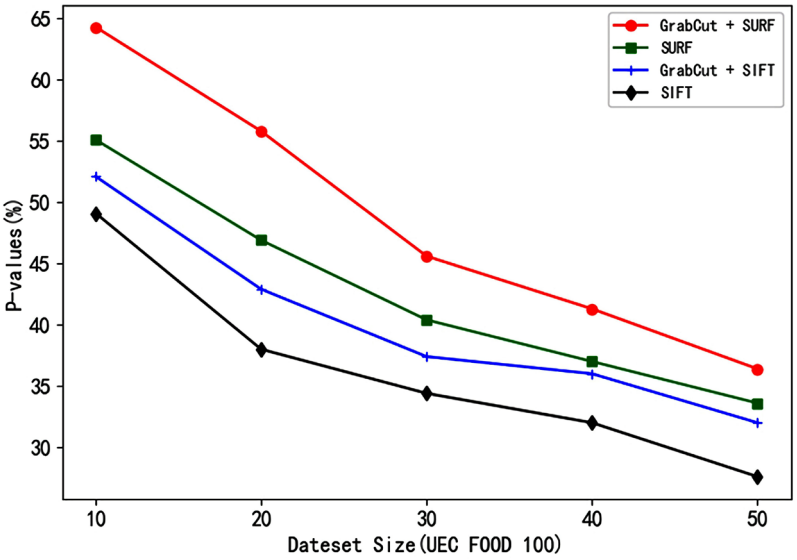

Figure 12. Comparison of experimental results before and afterimage segmentation.

Table 2. The effect of image segmentation on feature extraction.

Dataset size	Value of the neighbourhood radius	Image segmentation	Time (min)
10 types of images, 100 for each	$R = 2$	Yes	83.9
10 types of images, 100 for each	$R = 2$	No	149.8
20 types of images, 100 for each	$R = 4$	Yes	189.2
20 types of images, 100 for each	$R = 4$	No	343.2

4.3. Local Feature Fusion

The purpose of the experiment is to evaluate the influence of different value of the neighbourhood radius of feature points on the effect of local feature extraction. Therefore, we ignore the SURF information intentionally and extract only the colour information in the neighbourhood of the feature point as the final representation vector of the image. The evaluation criteria are the accuracy of the image classification, and **Figure 13** shows the experimental results.

The abscissa in **Figure 13** represents the value of the neighbourhood radius of the feature point, and the ordinate represents the accuracy of the image classification. The experiments were carried out on Caltech101 and UEC FOOD 100 respectively, took the top 30 types of images from Caltech101 and the first 20 types of images from UEC FOOD 100 for experiments. In the experiment, the colour information of all pixels in different value of the neighbourhood radius of SURF feature points is extracted and represented by RGB and HSV colour space respectively. Using the above information as the final representation vector of the image, then its dimension can be calculated from Equation (5). The final representation vector is quantized and input into the SVM for classification.

Analysis of the experimental results shown in **Figure 13**, we can see that as the value of the neighbourhood radius R continues to increase, the accuracy of image classification is also steadily increasing. After several rounds of testing, we can determine that when R is 5, we can get the best feature representation, and the RGB space is generally better for the representation of local colours.

At the same time, we found that the food image is more sensitive to local colour features than the integrated image. It is because the food images, especially Asian food images, their feature points are mostly located in the "embellishment area", and the colour information in the neighbourhood of the feature points has a higher value.

4.4. Global Feature Fusion

The purpose of this experiment is to evaluate the influence of global feature fusion on the overall feature extraction effect. Therefore, the single feature extraction method (SIFT/SURF) is used as a benchmark to compare the classification performance of local feature fusion (SIFT/SURF + RGB) and local-global feature fusion (SIFT/SURF + RGB + HSV). The evaluation criteria are the accuracy of the image classification, and **Figure 14** shows the experimental results.

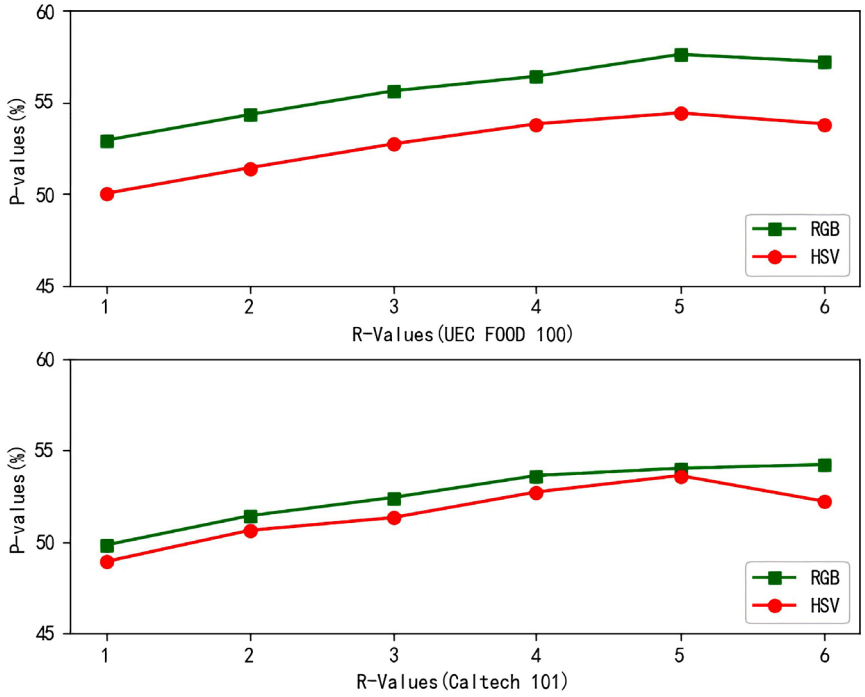

Figure 13. The effect of local colour features on classification.

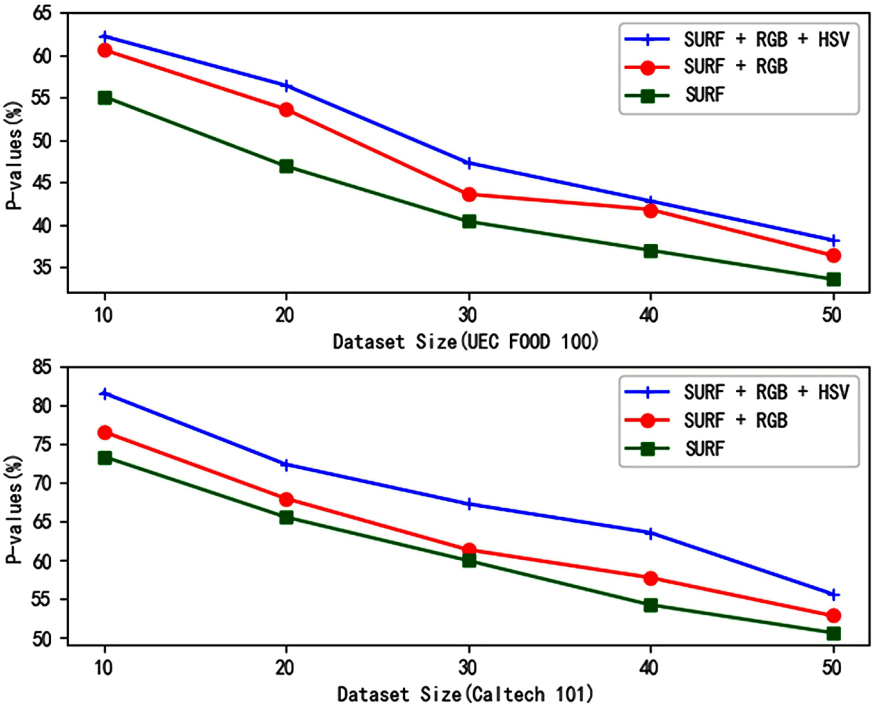

Figure 14. The effect of global color features on classification.

The abscissa in Figure 14 indicates the size of different datasets (the number of image types), and the ordinate indicates the classification accuracy. The experiments were carried out on Caltech101 and UEC FOOD 100, respectively, and the scale of the experimental data is consistent with the previous section. In the

experiment, we represent the overall colour information of the image in the HSV colour space and splicing it with the local features of the image (SIFT/SURF + RGB), wherein the value of the neighbourhood radius R takes a value of 2. Through the above comparison experiments, we found that the addition of the global colour feature improves the classification accuracy of the integrated image by about 5%, and contributes about 3% to the classification of food image. The global feature fusion, based on local feature fusion, further enhances the effect of feature extraction. The reason why global colour feature contributes less to the classification of food image is that the colour difference between the types of food image is smaller than that of the integrated image, and there are many types of foods having the same colour tone.

4.5. Experimental Results

We randomly selected 20 types of images in the UEC FOOD 100 as experimental data. According to Equation (4), the clustering value K is 1368, the value of the neighbourhood radius R is 5. After the local feature fusion, the dimension D_l of the local fused vector of each visual word is 247, according to the Equation (6). After the global feature fusion, the dimension D_g of the final representation vector of each image is 1440, according to the Equation (9). Under the above experimental conditions, we tested the components of the SLGC. The purpose is to summarize and quantify the contribution of image segmentation, local feature fusion and global feature fusion to the accuracy of image classification.

As can be seen from Table 3, the use of the SURF descriptor is better than the use of the SIFT descriptor, when the other conditions are the same, can generally improve the classification accuracy by about 4%. The segmentation of the food image can improve the classification accuracy by about 4% based on the same feature extraction method, the accuracy of classification can be improved by about 5% after local feature fusion, and 3% after the global feature fusion.

Then, using the same dataset and parameters as the above experiment, the SLGC is compared with other models (Table 4).

First, the baseline method Color Histogram and Bag of SIFT Features have a single feature extraction method, can not extract features according to the characteristics of food images, and reduced the effect of classification. OM performed well on the PFID, but because of its unique feature combination structure that

Table 3. Comparison of P values of different feature extraction methods.

Extraction methods	Image segmentation	Local feature fusion	Global feature fusion	P-Value/%
10 types of images, 100 for each	$R = 2$	Yes	83.9	48%/52%
10 types of images, 100 for each	$R = 2$	No	149.8	50%/56%
20 types of images, 100 for each	$R = 4$	Yes	189.2	57%/61%
20 types of images, 100 for each	$R = 4$	No	343.2	60%/64%

Table 4. Comparison of P values of different classification model.

Classification model	Dataset	Image content	P-Value/%
Color Histogram	UEC FOOD 100	Asian food	52%
Bag of SIFT Features	UEC FOOD 100	Asian food	49%
OM	PFID	Fast food	78%
OM	UEC FOOD 100	Asian food	57%
Texture + SIFT + MKL	UEC FOOD 100	Asian food	61%
SLGC	UEC FOOD 100	Asian food	64%

only applies to PFID, the accuracy of classification in Asian food datasets has declined.

On the UEC FOOD 100 dataset, we achieved similar performance to the Texture + SIFT + MKL, reaching more than 60%. At the same time, for the problem that it does not preprocess the original image, the GrabCut algorithm is used to extract the food subject from the image, which further improves the accuracy of classification, reaching about 64%.

5. Conclusions

To explore the intrinsic characteristics of Asian food images and improve the accuracy of classification, after studying and analyzing the unique structure and colour characteristics of Asian foods, this paper proposes an image classification model SLGC based on feature fusion. It constructs a feature representation method that combines SURF features, local colour information and global colour information, which can extract features of Asian food image comprehensively and efficiently. At the same time, the image segmentation algorithm is used to separate the invalid interference information and highlight the food subject, which further improves the effect of image classification. The experimental results show that the SLGC based on feature fusion can effectively improve the effect of image classification.

In the process of research, we have tried many feature matching and fusion schemes. The theoretical and practical basis of these schemes is not only computer vision related technologies, but also an in-depth study of the characteristics of Asian food. The research on the characteristics of Asian food in this article is not enough. In the future, we can try to fuse deeper texture and structural features to improve the understanding of Asian food pictures from the intensity of feature expression.

Acknowledgements

This work was supported by the Outstanding Youth Innovation Teams in Higher Education of Shandong Province (2019KJN048).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Subhi, M.A. and Ali, S.M. (2018) A Deep Convolutional Neural Network for Food Detection and Recognition. 2018 *IEEE-EMBS Conference on Biomedical Engineering and Sciences* (*IECBES*), Borneo, 3-6 December 2018, 284-287. https://doi.org/10.1109/IECBES.2018.8626720

[2] Tiankaew, U., Chunpongthong, P. and Mettanant, V. (2018) A Food Photography App with Image Recognition for Thai Food. 2018 *Seventh ICT International Student Project Conference* (*ICT-ISPC*), Nakhon Pathom, 11-13 July 2018, 1-6. https://doi.org/10.1109/ICT-ISPC.2018.8523925

[3] He, H., Kong F. and Tan, J. (2015) Dietcam: Multiview Food Recognition Using a Multikernelsvm. *IEEE Journal of Biomedical and Health Informatics*, **20**, 848-855. https://doi.org/10.1109/JBHI.2015.2419251

[4] Sasano, S., Han, X.H. and Chen, Y.W. (2016) Food Recognition by Combined Bags of Color Features and Texture Features. 2016 9*th International Congress on Image and Signal Processing*, *BioMedical Engineering and Informatics* (*CISP-BMEI*), Datong, China, 15-17 October 2016, 815-819. https://doi.org/10.1109/CISP-BMEI.2016.7852822

[5] Liang, G., Hong, H., Xie, W., *et al.* (2018) Combining Convolutional Neural Network with Recursive Neural Network for Blood Cell Image Classification. *IEEE Access*, **6**, 36188-36197. https://doi.org/10.1109/ACCESS.2018.2846685

[6] Zhao, W., Jiao, L., Ma, W., *et al.* (2017) Superpixel-Based Multiple Local CNN for Panchromatic and Multispectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 4141-4156. https://doi.org/10.1109/TGRS.2017.2689018

[7] Hnoohom, N. and Yuenyong, S. (2018) Thai Fast Food Image Classification Using Deep Learning. 2018 *International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering* (*ECTI-NCON*), 116-119. https://doi.org/10.1109/ECTI-NCON.2018.8378293

[8] Ege, T. and Yanai, K. (2017) Estimating Food Calories for Multiple-Dish Food Photos. 2017 4*th IAPR Asian Conference on Pattern Recognition* (*ACPR*), Nanjing, 26-29 November 2017, 646-651. https://doi.org/10.1109/ACPR.2017.145

[9] Naritomi, S., Tanno, R., Ege, T. and Yanai, K. (2018) FoodChangeLens: CNN-Based Food Transformation on HoloLens. 2018 *IEEE International Conference on Artificial Intelligence and Virtual Reality* (*AIVR*), Taichung, Taiwan, 10-12 December 2018, 197-199. https://doi.org/10.1109/AIVR.2018.00046

[10] Sharma, O. (2019) Deep Challenges Associated with Deep Learning. 2019 *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing* (*COMITCon*), Faridabad, India, 14-16 February 2019, 72-75. https://doi.org/10.1109/COMITCon.2019.8862453

[11] Wang, D.Y. (2019) Comparison and Reference of the Differences between Chinese and Western Food and Nutrition Development. *Food and Nutrition in China*, **25**, 5-8.

[12] Jaisakthi, S.M., Mirunalini, P. and Aravindan, C. (2018) Automated Skin Lesion Segmentation of Dermoscopic Images Using GrabCut and k-Means Algorithms. *IET Computer Vision*, **12**, 1088-1095. https://doi.org/10.1049/iet-cvi.2018.5289

[13] Ren, D., Jia, Z., Yang, J., *et al.* (2017) A Practical Grabcut Color Image Segmentation Based on Bayes Classification and Simple Linear Iterative Clustering. *IEEE Access*, **5**, 18480-18487. https://doi.org/10.1109/ACCESS.2017.2752221

[14] Mustafa, R. and Dhar, P. (2018) A Method to Recognize Food Using Gist and SURF Features. 2018 *Joint 7th International Conference on Informatics, Electronics & Vision* (*ICIEV*) *and* 2018 2*nd International Conference on Imaging, Vision & Pattern Recognition* (*icIVPR*), 127-130. https://doi.org/10.1109/ICIEV.2018.8641072

[15] Zhu, Q., Zhong, Y., Zhao, B., *et al.* (2016) Bag-of-Visual-Words Scene Classifier with Local and Global Features for High Spatial Resolution Remote Sensing Imagery. *IEEE Geoscience and Remote Sensing Letters*, **13**, 747-751. https://doi.org/10.1109/LGRS.2015.2513443