Research Article

Multiview Machine Vision Research of Fruits Boxes Handling Robot Based on the Improved 2D Kernel Principal Component Analysis Network

Xinning Li [ID],[1] **Hu Wu,**[1] **Xianhai Yang** [ID],[1] **Peng Xue,**[2] **and Shuai Tan**[2]

[1]*School of Mechanical Engineering, Shandong University of Technology, Zibo 255000, China*
[2]*National Engineering Research Center for Production Equipment, Dongying 257091, China*

Correspondence should be addressed to Xianhai Yang; yxh@sdut.edu.cn

In order to better realize the orchard intelligent mechanization and reduce the labour intensity of workers, the study of intelligent fruit boxes handling robot is necessary. The first condition to realize intelligence is the fruit boxes recognition, which is the research content of this paper. The method of multiview two-dimensional (2D) recognition was adopted. A multiview dataset for fruits boxes was built. For the sake of the structure of the original image, the model of binary multiview 2D kernel principal component analysis network (BM2DKPCANet) was established to reduce the data redundancy and increase the correlation between the views. The method of multiview recognition for the fruits boxes was proposed combining BM2DKPCANet with the support vector machine (SVM) classifier. The performance was verified by comparing with principal component analysis network (PCANet), 2D principal component analysis network (2DPCANet), kernel principal component analysis network (KPCANet), and binary multiview kernel principal component analysis network (BMKPCANet) in terms of recognition rate and time consumption. The experimental results show that the recognition rate of the method is 11.84% higher than the mean value of PCANet though it needs more time. Compared with the mean value of KPCANet, the recognition rate exceeded 2.485%, and the time saved was 24.5%. The model can meet the requirements of fruits boxes handling robot.

1. Introduction

As the primary industry of the national economy, agriculture is the primary condition for all production, and the proposal of precision agriculture has put forward higher requirements. The fruit industry, as a labour-intensive industry, has a large demand for labour and low work efficiency. The automation and mechanization industry chain needs upgrade urgently [1]. With the rapid development of the artificial intelligence, the fruit recognition and fruit picking have always being studied [2]. There are relatively few studies on fruit handling [3]. On farms and in wholesale fruit markets, the handling of fruits boxes is still dominated by manual labour, which is time-consuming and labour-consuming. In the new era, the cost of manual labour is increasing year by year, which cannot meet the demand of precision agriculture. In order to realize intelligent handling, this paper studied the fruits boxes recognition based on machine vision.

According to the different modelling methods of target appearance, the research results of target recognition in recent years have been divided into three categories [4]: based on feature invariants, representation learning, and deep learning. The view models based on feature invariants extract the features of multiple images from different perspectives and then train the classifier, which are used for the occasions with a small number of training samples. The research studies mainly focus on the construction of artificial features and classification algorithms, and many outstanding works have emerged. Due to the necessity to study the characteristic invariance of the target in advance, candidate features have characters such as weak adaptability, weak

generalization ability, and large application limitation. It has the large feature description vector dimension and high training cost of the classifier. Researchers proposed the methods based on subspace learning to solve the problems, which transformed high-dimensional feature vectors into low-dimensional ones. The classifiers were trained in the subspace. The typical representative methods are as follows: principal component analysis (PCA) based on unsupervised learning and linear discriminant analysis (LDA) based on supervised learning. Based on these, the methods with low data dimension, strong noise processing ability, and high efficiency were put forward, such as robust PCA (RPCA), inductive RPCA (IRPCA), kernel PCA (KPCA), two-dimensional PCA (2DPCA), and discriminative low-rank and sparse principal feature coding (D-LSPFC). With the emergence of a large number of public image datasets, the target recognition methods based on deep learning have been studied more and more. The models based on convolutional neural network (CNN) promoted the development of computer vision in particular by virtue of its strong nonlinear feature expression ability and good generalization performance. Region CNN (R-CNN) [5] applied deep learning in the target recognition for the first time. And then deep convolutional neural networks Fast R-CNN [6] and Faster R-CNN [7] were proposed by combining the training and testing process, which improved the identification accuracy and efficiency greatly. As the product of integrating fuzzy logic reasoning and self-learning ability of neural network, neurofuzzy network has also been widely used [8]. The CNN-based single shot detector (SSD) [9] and the YOLO [10] deep learning object detection method further improved a new height in real-time effect. On this basis, the proposed YOLOv2 [11] and YOLO V3 [12] gradually improved the running speed and robustness, and the detection performance had been significantly improved. YOLO V4 [13] achieved double improvement in speed and accuracy, which took CSPDarkNet53 + SPP + PANet (path-aggregation-neck) + yolov3-head as the model. It is undeniable that the effect of the target recognition algorithm based on deep learning is remarkable, and the recognition accuracy is much higher than the traditional manual methods and the representation learning methods. However, it cannot be ignored that the target recognition still has great challenges in some occasions, such as target overlap, partial occlusion, high similarity, complex environment, and strong interference. The methods with complex models, long training time, and high requirements for hardware computing power have affected the application in mobile robots.

As a three-dimensional (3D) object, the direct extraction and recognition of 3D features for fruit box lead to complex calculation and high operation storage. A view-based method is adopted in this paper, that is, 3D objects are represented through multiple views. As a common method of 3D objects recognition, multiview learning model and recognition method have also received more attention. The multiview-based convolutional neural network (MVCNN) was built in [14]. The maximum pooling layers blended the multiple views features. The MVCNN model had low convergence speed and low

efficiency of feature extraction because it was not end-to-end network. The end-to-end group-pair convolutional neural network (GPCNN) was established in [15]. The small-scale problem could be solved. The novel pairwise multiview convolutional neural network (PMV-CNN) was proposed in [16], which focused on complementary information between views. The feature extraction and target recognition are unified into CNN. It could improve the robustness of feature extraction obviously when the number of training samples was small. In order to make up for the disadvantages caused by random images selection in multiview recognition, a multiview discrimination and pairwise convolutional neural network (MDPCNN) was obtained by adding the Slice layer and the Concat layer in [17]. The model was verified that it had good intraclass compactness and interclass separability. The multiview-based Siamese convolutional neural network was exploited in [18]. An end-to-end multiview 3D fingerprint learning model was proposed in [19], which included full convolutional network and three Siamese networks. The multiview generator module was used in [20] to project the 3D point cloud to the plane at a specific angle. On the premise of retaining the underlying features, spatial refusion operation was adopted to realize the interaction between different projections, and the features were reconstructed for target recognition. Based on the semi-supervised learning and expectation maximization, a multiview fusion strategy classification method with the ability of label propagation was proposed in [21]. An end-to-end cloud convolutional neural network was built in [22] based on the projection network mechanism. The point cloud was projected into a two-dimensional view with rich discriminant features, and the robustness and accuracy had been improved significantly. Multiview features projections were coded as binary in [23]. The recognition descriptors were assembled block statistical features. Although the above methods have achieved good results, the models based on the convolutional neural network (CNN) also have some problems, such as complicated structure, long training cycle. They do not seem to be the best choice for fruits boxes handling robots. From the above research, it can be concluded that considering the relationship between multiple views can improve the recognition accuracy and robustness, and the binary coding method can improve the operation efficiency of the model, which also become the research factors of the new model developed in this paper.

No system is perfect. The hidden state of the inevitable uncertainty in the system can be stimulated, and the connection between these uncertainties and the object system can be established to improve the system performance [24]. Although fruit packing boxes are generally regular cubes, traditional rule-based feature extraction and recognition methods cannot achieve better results because of the variety of fruits and the influence of surface patterns, colours, and surrounding environment. Therefore, deep learning algorithm is more advantageous. The current deep learning target recognition algorithm is an end-to-end solution; that is, it is completed in one step from the

input image to the output task result, but it is completed in stages internally as image feature extraction network classification and regression. Aiming at the long training time of the classic CNN parameters, the simple principal component analysis network (PCANet) was built in [25]. The convolution layer of CNN was introduced into the classical feature extraction framework of "Feature Map-Pattern Map-Histogram." The unsupervised hierarchical features were obtained. The high computational complexity caused by iteration and optimization was avoided. It has been widely used with simple model and rapid calculation. Since PCA could not extract the nonlinear relationship between images, the kernel principal component analysis network (KPCANet) model was established in [26], which achieved better classification results than PCANet. In order to remove the redundancy of multiperspective views, our team proposed a binary multiview kernel principal component analysis network (BMKPCANet) model [27] for the multiview objects recognition. However, the model converted two-dimensional image matrix into vector when the features were extracted, and the original image structure was destroyed, and the computation was also large, so we improved the model. Inspired by the two-dimensional principal component analysis network (2DPCANet) [28] and the two-dimensional kernel principal component analysis (2DKPCA) [29], the images of fruits boxes were processed by 2DKPCA, and a new multiview feature extraction model was established. The main contributions of this work were summarized as follows:

(1) A binary multiview two-dimensional kernel principal component analysis network (BM2DKPCANet) model was built to extract clustering features, which can reduce data redundancy and realize binary multiview clustering.

(2) The multiview recognition method of fruits boxes was proposed combining BM2DKPCANet model with the support vector machine (SVM).

(3) The proposed method was compared with the PCANet, 2DPCANet, KPCANet, and BMKPCANet models on the fruits boxes dataset and ETH-80 and COIL-100 public datasets. Taking the recognition accuracy and time consumption as the evaluation indexes, the experiments showed that the recognition performance of the proposed method was superior to other methods.

The rest of this work was organized as follows. The methods based on the 2DPCA and KPCA are introduced in Section 2. The obtaining method of fruits boxes images from multiview angles is introduced in Section 3. The feature extraction process of the proposed BM2DKPCANet algorithm and the identification process of fruits boxes are also discussed in detail in Section 3. The experimental process, results, and discussion are shown in Section 4. Finally, the research and the future work are summarized in Section 5.

2. Related Works

In order to reduce the sample dimension and obtain the nonlinear correlation between multiple pixels, some scholars have proposed a series of algorithms by synthesizing the advantages of 2DPCA [30] and KPCA [31]. Nhat and Lee [32] proposed the kernel-based 2DPCA, which directly extracted nonlinear features from two-dimensional images. The nonlinear correlation analysis of matrix was realized. However, the storage requirement of kernel matrix was higher when training samples were large. Zhou et al. [33] calculated the low-rank approximate decomposition of kernel matrix using Cholesky decomposition method to achieve nonlinear feature extraction. The computational efficiency was low in the test stage. Xu et al. [34] used Laplace to reduce dimension after the 2DKPCA. Choi et al. [35] proposed the incremental 2DKPCA (I2DKPCA), which reduced the calculation speed and improved the performance of feature extraction. Zhang et al. [29] built the 2-dimensional kernel PCA (2DKPCA) framework. The performance of unilateral 2DKPCA (row and column) and that of bilateral 2DKPCA in face and object recognitions were compared, respectively. Mohammad et al. [36] matched historical parameters by bilateral 2DKPCA. Xiang et al. [37] realized dimensionality reduction for hyperspectral images using the segmented row-column K2DPCA method. In order to reduce the storage requirement and computational complexity of kernel matrix, blockwise methods were proposed [38, 39], which transformed the large kernel matrix into several small kernel matrices and then combined the eigenvectors of small kernel matrices. Wang and Zhou [40] mixed image blocks and vector method. The scale of the kernel matrix was decreased by taking several adjacent rows or columns of the graph as a computing unit for non-mapping. Chen et al. [41] proposed bidirectional two-dimensional kernel quaternion principal component analysis (BD2DKQPCA) for colour image recognition. The kernel matrix was used to replace the covariance matrix between samples, which avoided the high-complexity calculation of high-dimensional space. Then they improved 2DKQPCA by adding blockwise process [42]. According to the characteristics of quaternion Hermitian matrix, the blocks of main diagonal, next to the main diagonal, and backdiagonal direction were analyzed.

Through the research of the above algorithms, considering the recognition and computing performance, this paper sampled images in blocks when extracting the image features. It had been demonstrated that the recognition performance of the column-oriented algorithm was superior to the row-oriented algorithm by experiments in the proposed B2DKPCA [38], the bidirectional two-dimensional KQPCA (BD2DKQPCA) [41], and the block-based 2DKQPCA [42]. So this work adopted column-oriented algorithm to conduct 2DKPCA; that is, the column vector of the image sample is mapped to a high-dimensional space through the nonlinear mapping function. The kernel matrix replaced the covariance matrix.

3. Materials and Methods

3.1. Experimental Materials

3.1.1. Establishment of Multiview Dataset of Fruits Boxes. This work adopted the multiview feature method to collect images. Under the principle of ensuring that the set of projected views is as small as possible and can represent many common attitudes of the boxes, several two-dimensional projections with different viewpoints are used to describe the features of the boxes. In order to describe and establish visual model preferably, the relative position relation between fixed view and boxes in different positions was transformed into the relation between relative movement view and fixed boxes. Various observed postures of the boxes in normal operation were collected under the motion view. Since the opposite sides of the boxes had the same pattern generally, multiple semiarc viewpoint projection model was set up, as shown in Figure 1. The camera kept moving on the green cambered surface, and the multiple different postures of the boxes are obtained.

The semiarc viewpoints surface must be divided into small areas to obtain the projection of 3D targets with different attitudes. The view areas are reasonably divided and distributed viewpoints to ensure that the projection view set is as small as possible and can represent multiple common attitudes of the boxes. The distribution of viewpoints was described by the representation of latitude and longitude in geography based on the idea of uniform division and morphology diagram method [43] to simulate the box postures in the real situation, as shown in Figure 2. The projection of the box at each viewpoint corresponds to a two-dimensional image, respectively, and the multiview projection model of the box was constructed.

The experimental objects were from the fruit wholesale market of Zhangdian District, Zibo City, Shandong Province, China. A total of 15 different types for 10 kinds of fruits boxes were selected in the experiment, which were defined as apple1, apple2, apple3, watermelon1, watermelon2, orange1, orange2, cantaloupe1, cantaloupe2, pomegranate, pear, durian, coconut, banana, and pineapple, as shown in Figure 3. Multiview collection was carried out for the boxes of each category, which is shown in Figure 4. 200 samples of each category were retained. The image size was normalized to 32×32, and gray processing was carried out.

3.1.2. Multiview Public Datasets. In order to fully verify the performance of the proposed multiview recognition algorithm, the recognition performance tests are also carried out on public datasets ETH-80 [44] and COIL-100 [45]. The ETH-80 dataset contains 8 species classes. Each species is an image set of 10 different objects, which contains 41 images of each object taken from different angles. 4 objects of each species were randomly selected to form the training set, and the rest were used as the test set in this paper. The COIL-100 dataset contains images of 100 objects. Each object was taken at 72 different angles within a 360° circumference. The training set was
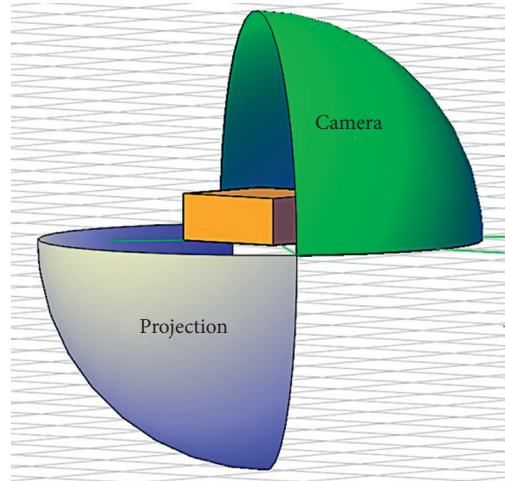

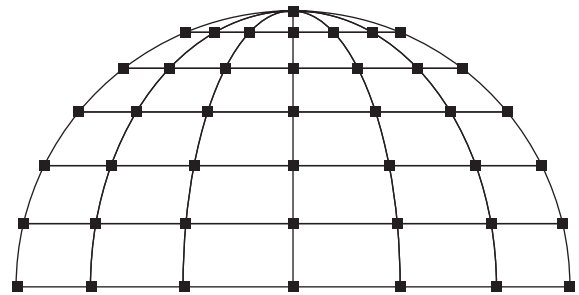FIGURE 1: Image acquisition method.

FIGURE 2: Distribution of viewpoints.

composed of the 720 images of 20 objects randomly. The partial images of the ETH-80 and COIL-100 are shown in Figure 5.

3.2. BM2DKPCANet Model Based on 2DKPCA

3.2.1. Construction of Feature Extraction Model. Since the image database is composed of several two-dimensional multiview images, the images as much as possible represent the common postures of the boxes, which lead to a lot of data redundancy. In order to reduce unnecessary data storage, this paper added clustering step in the feature extraction model of fruits boxes, as shown in Figure 6. According to the related research principal component analysis network, the two-layer 2DKPCA network was constructed. The extracted feature vectors were binary clustering coded at the same time. The clustering feature representation of decimal system was obtained by block histogram transformation, and the clustering feature extraction was completed.

3.2.2. BM2DKPCANet Model

(1) First 2DKPCA. The image size of database was adjusted to $m \times n$. As the input layer I_i, patch sampling was sliding performed by $k \times k$ window. All sample patches were gathered and cascaded. The jth patch of the ith image was defined as $x_{i,j}$. The ith image could be expressed as

FIGURE 3: Categories of fruits boxes dataset.

FIGURE 4: Collecting images in multiviews.

$$\widehat{\mathbf{I}}_i = \left[\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \ldots, \mathbf{x}_{i,\widehat{m}\cdot\widehat{n}}\right], \tag{1}$$

where $\widehat{m}$ was the number of patches on rows and $\widehat{n}$ was the number of patches on columns. The demean sample patch was obtained as follows:

$$\overline{\mathbf{x}_{i,j}} = \mathbf{x}_{i,j} - \frac{\sum_{j=1}^{\widehat{m}\cdot\widehat{n}} \mathbf{x}_{i,j}}{\widehat{m}\cdot\widehat{n}}. \tag{2}$$

The local feature matrix of the ith image could be written as

$$\mathbf{X}_i = \left[\overline{\mathbf{x}_{i,1}}, \overline{\mathbf{x}_{i,2}}, \ldots, \overline{\mathbf{x}_{i,\widehat{m}\cdot\widehat{n}}}\right] \in R^{k^2 \times mn}. \tag{3}$$

After doing the same progress for the other images, the feature analysis based on 2DKPCA was performed on the local feature matrix. Due to not needing explicit form after mapping, and in order to avoid complex calculation in high-dimensional space, the covariance matrix after samples mapping was replaced by kernel matrix [41]. Training sample matrix $\mathbf{I}$ i ($i = 1, 2, \ldots, S$) $\in R^{m \times n}$ was converted to local eigenmatrix $\mathbf{X}_i$ ($i = 1, 2, \ldots, S$) $\in R^{k^2 \times mn}$ after patches sliding sampling. The dimension of the column direction kernel matrix for S training samples is $Smn \times Smn$, which requires a large amount of computation. This work adopted average column vectors to replace the original mn column vectors [29]; then the sample of nonlinear mappings for training became $\overline{\mathbf{X}}_i$; that is,

$$\Psi: R^{k^2} \longrightarrow F,$$
$$\mathbf{X}_i^t \longrightarrow \phi\left(\overline{\mathbf{X}}_i\right). \tag{4}$$

$$\overline{\mathbf{X}}_i = \frac{1}{mn}\sum_{t=1}^{mn}\mathbf{X}_i^t, \tag{5}$$

where $i = 1, 2, \ldots, S$, $t = 1, 2, \ldots, mn$. S training samples can be approximately expressed in the kernel feature space as follows:

$$\Phi_1 = \left[\phi\left(\overline{\mathbf{X}}_1\right), \phi\left(\overline{\mathbf{X}}_2\right), \ldots, \phi\left(\overline{\mathbf{X}}_S\right)\right]. \tag{6}$$

Then the kernel matrix can be expressed as

$$\mathbf{K}_1 = \Phi_1^T \Phi_1 = \left[k\left(\overline{\mathbf{X}}_s, \overline{\mathbf{X}}_s\right)\right] = \left[\langle\phi\left(\overline{\mathbf{X}}_s\right), \phi\left(\overline{\mathbf{X}}_s\right)\rangle\right]. \tag{7}$$

The dimension was reduced to $S \times S$, and the computational complexity reduced greatly. The kernel matrix $K1$ was centralized [46], such that

$$\overline{\mathbf{K}}_1 = \mathbf{K}_1 - \frac{1}{S}\mathbf{K}_1\mathbf{1} - \frac{1}{S}\mathbf{1}\mathbf{K}_1 + \frac{1}{S^2}\mathbf{1}\mathbf{K}_1\mathbf{1}, \tag{8}$$

(a)   (b)
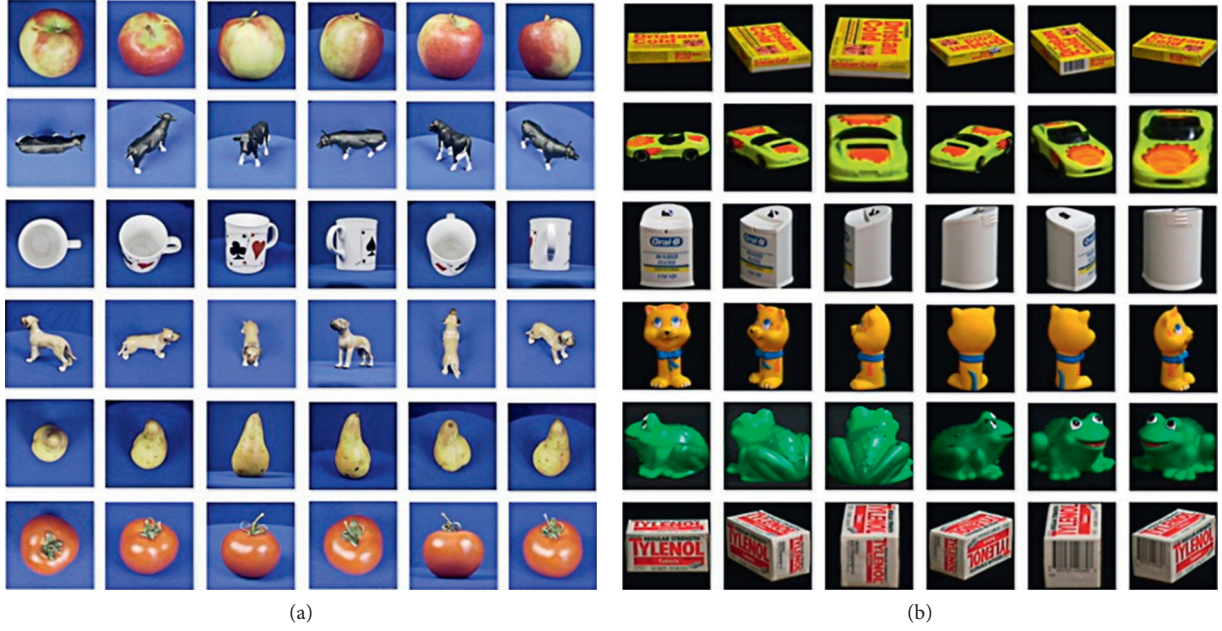
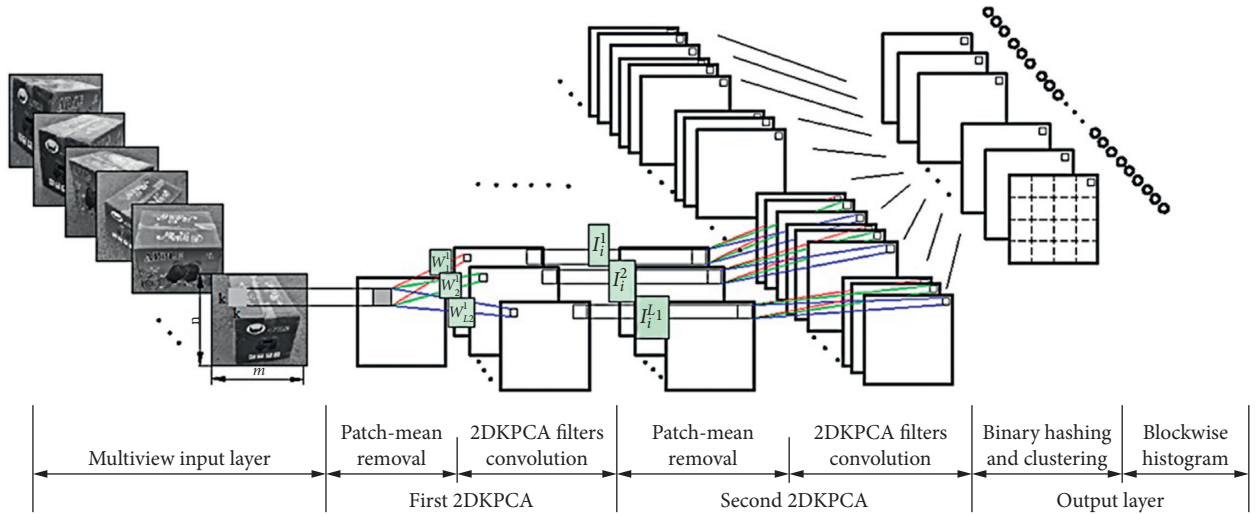FIGURE 5: Partial samples of the public datasets (a) ETH-80 and (b) COIL-100.

FIGURE 6: Model of BM2DKPCANet.

where $\mathbf{1}$ was the matrix of order S whose all components were 1. The eigenvectors corresponding to the top L_1 largest eigenvalues of $\overline{\mathbf{K}}_1$ were taken as the kernel principal component filters of the first-layer network.

$$\mathbf{W}_l^1 = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{L_1}\right]. \tag{9}$$

The training sample $\mathbf{I}_i$ after the zero-filled boundary was convolved with the first-layer 2DKPCA filter,

$$\mathrm{I}_i^l = \mathrm{I}_i * \mathbf{W}_l^1, i = 1, 2, \ldots, S; l = 1, 2, \ldots, L_1, \tag{10}$$

where $*$ was two-dimensional convolution, $\mathrm{I}_i^l \in \mathrm{R}^{m \times n}$, and L_1 was the filters number of the first 2DKPCA.

(2) Second 2DKPCA. Taking the output of the first 2DKPCA as the input of the second 2DKPCA, the same process as the first 2DKPCA was repeated. The nonlinear high-dimensional mapping of the image matrix was carried out. The kernel matrix $\mathbf{K}_2$ was calculated and centralized to $\overline{\mathbf{K}}_2$ approximately. The first L_2 kernel principal component features of $\overline{\mathbf{K}}_2$ were used as the filters convolution kernel $\mathbf{W}_\ell^2$ of the second-layer network:

$$\mathbf{W}_\ell^2 = \left[\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{L_2}\right]. \tag{11}$$

Similarly, the output of the first 2DKPCA was further convoluted, and the output of the second 2DKPCA could be obtained:

$$O_i^l = I_i^l * \mathbf{W}_\ell^2 = {}_1^i * \mathbf{W}_l^1 * \mathbf{W}_\ell^2, \tag{12}$$
$$\text{s. t. } i = 1, 2, \ldots, S; l = 1, 2, \ldots, L_1; \ell = 1, 2, \ldots, L_2.$$

(3) Binary Hash Features Clustering. Similarly, the output of the first 2DKPCA was further convoluted, and the output of the second 2DKPCA could be obtained: in order to reduce the data redundancy caused by the multiangle acquisition process of the box image, the clustering operation was carried out in this stage. Binary clustering algorithm [47]

used binary encoding technology to solve the problem of multiview clustering. Binary encoding and clustering for multiple views were jointly optimized at the same time. The problems of big data storage and long time-consuming operation were well improved. It reduced the computation time and storage space greatly. The speed and efficiency were enhanced. The model proposed in this paper encoded and clustered multiview dataset at the same time, and the total optimization function was set as

$$\min F(\mathbf{U}^m, \mathbf{B}, \mathbf{C}, \mathbf{G}, \mathbf{a}) = \sum_{m=1}^{M} (\mathbf{a}^m)^r \left(\left\| \mathbf{B} - \mathbf{U}^m \phi(O^{lm}) \right\|_F^2 + \beta \|\mathbf{U}^m\|_F^2 \right.$$
$$\left. \cdot \left(-\frac{\gamma}{n} tr \left(\mathbf{U}^m \phi((O^{lm})) (\mathbf{U}^m \phi(O^{lm}))^T \right) \right) + \lambda \|\mathbf{B} - \mathbf{CG}\|_F^2 \right. \tag{13}$$
$$\text{s.t. } C^T 1 = 0, \sum_m \boldsymbol{\alpha}^m = 1, \boldsymbol{\alpha}^m > 0, \mathbf{B} \in \{-1, 1\}^{q \times n}, \mathbf{C} \in \{-1, 1\}^{q \times c}, \mathbf{G} \in \{0, 1\}^{c \times n}, \sum_j \mathbf{g}_{ji} = 1,$$

where α^m was the weight of the mth view, $m = 1, \ldots, M$. Different views had different weights. $r > 1$ was scalar that controlled the weights. $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$, $\mathbf{b}_i$ was collaborative binary code of the ith instance, and each encoding B was represented by the product of a clustering centroid C and indicator vector g. $\mathbf{U}^m$ was mapping matrix of mth view. $\phi(O^{lm})$ was the kernel function based on nonlinear RBF mapping between the output feature of the second 2DKPCA and selected sample points randomly under the mth view. γ

was nonnegative constant. $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_n]$ and λ are the regularization parameters.

The optimization problem was divided into several subproblems. $\mathbf{U}^m$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{G}$, and α^m were optimized and updated alternately by an alternating optimization strategy. When some variable was updating, other variables were fixed. The corresponding optimization cost functions were as follows; then the sample of nonlinear mappings for training became $\overline{\mathbf{X}}_i$; that is,

$$\min F(\mathbf{U}^m) = \left\| \mathbf{B} - \mathbf{U}^m \phi(O^{lm}) \right\|_F^2 + \beta \|\mathbf{U}^m\|_F^2 - \frac{\gamma}{n} tr \left((\mathbf{U}^m \phi(O^{lm})) (\mathbf{U}^m \phi(O^{lm}))^T \right), \tag{14}$$

$$\min_{\mathbf{B}} \sum_{m=1}^{M} (\boldsymbol{\alpha}^m)^r \left(\left\| \mathbf{B} - \mathbf{U}^m \phi(O^{lm}) \right\|_F^2 \right) + \lambda \|\mathbf{B} - \mathbf{CG}\|_F^2$$
$$= tr \left[\mathbf{B}^T \left(\sum_{m=1}^{M} (\boldsymbol{\alpha}^m)^r \mathbf{I} + \lambda \mathbf{I} \right) \mathbf{B} \right] - 2tr \left[\mathbf{B}^T \left(\sum_{m=1}^{M} (\boldsymbol{\alpha}^m)^r \mathbf{U}^m \phi(O^{lm}) + \lambda \mathbf{CG} \right) \right] + \text{con}, \tag{15}$$

$$\min F(\mathbf{C}) = \|\mathbf{B} - \mathbf{CG}\|_F^2 + \rho \left\| \mathbf{C}^T \mathbf{1} \right\|^2 = -2tr(\mathbf{B}^T \mathbf{CG})) + \rho \left\| \mathbf{C}^T \mathbf{1} \right\|^2 + \text{con}, \tag{16}$$

$$\mathbf{g}_{js}^{p+1} = \begin{cases} 1, & s = \arg \min_i H(\mathbf{b}_j, \mathbf{c}_i^{p+1}), \\ 0, & \text{otherwise}, \end{cases} \tag{17}$$

$$\boldsymbol{\alpha}^m = \frac{(\mathbf{g}^m)^{1/(1-r)}}{\sum_m (\mathbf{g}^m)^{1/(1-r)}}, \tag{18}$$

where con is the constant with respect to B. H is the distance from each $\mathbf{B}$ to the cluster center. Until the total optimization function was optimal, the binary hash clustering optimization was completed.

(4) Output of the Block Histogram Features. L_2 features were outputted for each input I_i^l in second 2DKPCA, whose binary cell vector was clustered and optimized as a whole. Each optimized feature was converted to decimal,

$$\mathbf{T}_i^{lm} = \sum_{l=1}^{L_2} 2^{l-1}\left(\mathbf{h}_i^{lm}\right), \tag{19}$$

where $\mathbf{T}_i^{lm}, l \in [1, L_1]$, and each pixel was an integer within $[0, 2^{L_2-1}]$. Z blocks of each $\mathbf{T}_i^{lm}$ were counted by histogram $\mathrm{Zhist}(\mathbf{T}_i^{lm})$. A vector can be obtained by connecting $\mathrm{Zhist}(\mathbf{T}_i^{lm})$,

$$\mathbf{f}_i^m = \left[\mathrm{Zhist}\left(\mathbf{T}_i^{1m}\right), \cdots, \mathrm{Zhist}\left(\mathbf{T}_1^{Lm}\right) \right]^T \in \mathbb{R}^{(2^{L_2})L_1 Z}, \tag{20}$$

where $\mathbf{f}_i^m$ is the BM2DKPCANet feature of the ith sample under the mth view.

3.2.3. Fruits Boxes Recognition Based on BM2DKPCANET Model.
The fruits boxes features extracted by BM2DKPCANet model were input into the classifier for training recognition. The performance of classifier determines the recognition accuracy and classification speed directly. Support vector machine (SVM) is widely used in the field of pattern recognition because of its outstanding advantages in solving small sample nonlinear high-dimensional pattern recognition [48, 49]. This work also used SVM as classifier. According to previous studies [27], the radial basis function (RBF) was selected as kernel function, which mapped the features into the high-dimensional space to find the optimal hyperplane. Correct recognition of different kinds of fruits boxes achieved. The specific identification process of fruits boxes is shown in Figure 7.

4. Results and Discussion

The experiment was performed by Matlab2017b and Python integrated environment Anaconda3 on the Intel(R) Xeon(R) CPU E5-1650 v4@3.6 GHz, 64 GB RAM, NVIDIA GeForce GTX 10808G GPU platforms. The classifier kernel parameters were selected by grid search method and cross-validation method based on the LibSVM software package. The penalty parameters $C = 58$ and $\gamma = 2$ were determined. The following experiment analyzed influence of parameters on model performance taking the average accuracy after 10 tests as the evaluation index. The recognition accuracy was used as the evaluation metric:

$$\mathrm{accuracy} = \frac{\sum_{i=1}^{n} \phi\left(\hbar_i, \mathrm{map}\left(c_i\right)\right)}{n}, \tag{21}$$

where n is the total number of images in the dataset, g_i is the ground-truth of images, and $\mathrm{map}(c_i)$ is the classification predicted by the algorithm. If $\hbar_i = \mathrm{map}(c_i)$, then $\phi(\hbar_i, \mathrm{map}(c_i)) = 1$; otherwise $\phi(\hbar_i, \mathrm{map}(c_i)) = 0$.

4.1. Influence of Kernel Function.
KPCA is a nonlinear extension of principal component analysis using kernel technique. The selection of kernel function is related to the extraction of nonlinear features of dataset and affects the performance of model recognition directly. This paper studied the influence of commonly used kernel functions on the performance of the BM2DKPCANet model, such as

Linear, Polynomial, PolyPlus, Gaussian, and Sigmoid kernel functions. Their corresponding expressions are as follows:

$$K\left(\mathbf{v}_i, \mathbf{v}_j\right) = \mathbf{v}_i\mathbf{v}_j^T + c, \tag{22}$$

$$K\left(\mathbf{v}_i, \mathbf{v}_j\right) = \left(\mathbf{v}_i\mathbf{v}_j^T\right)^d, \tag{23}$$

$$K\left(\mathbf{v}_i, \mathbf{v}_j\right) = \left[\left(\mathbf{v}_i\mathbf{v}_j^T\right) + 1\right]^d, \tag{24}$$

$$K\left(\mathbf{v}_i, \mathbf{v}_j\right) = \exp\left(-\frac{\left\|\mathbf{v}_i - \mathbf{v}_j\right\|^2}{2\sigma^2}\right), \tag{25}$$

$$K\left(\mathbf{v}_i, \mathbf{v}_j\right) = \tanh\left(\alpha\mathbf{v}_i\mathbf{v}_j^T + c\right), \tag{26}$$

where c, d, σ, α are all real constants. This work defined that $c = 0, d = 3, \sigma = 1$, and $\alpha = 1/2$. $v_i, v_j \in \mathrm{R}^{mn}$ is the row vector of the matrix to be transformed. The influence of kernel function on model performance in the same parameters environment was studied, as shown in Figure 8. It can be seen that Gaussian kernel function adopted in the model can achieve the best recognition effect.

4.2. Influence of Filter Parameters

4.2.1. Influence of Number of Filters.
The patch size, block size, and overlapping ratio were set as 5×5, 8×8, and 0.5, respectively. The influence of the number of filters on the performance of the model was studied, as shown in Figure 9. The blue line represents the accuracy on the fruits boxes dataset when the number of filters of the first 2DKPCA was selected within range from 2 to 14. The accuracy tended to be stable when $L_1 \geq 8$. The selection of the second 2DKPCA filter was conducted with $L_1 = 8$. The red line represents the accuracy on the fruits boxes dataset when the number of filters of the second 2DKPCA was selected within range from 2 to 14. The accuracy is levelling off when $L_2 \geq 6$. $L_1 = 8, L_2 = 6$ would be set in the following experiment.

4.2.2. Influence of Patch Size and Block Size.
Since PCA filter has the conditions of $k_1 k_2 \geq L_1, L_2$, the minimum patch size was set to 3×3. In order to observe the influence of patch size and block size on the recognition in this proposed model, the block sizes were defined as 4×4, 8×8, and 16×16. The maximum patch size was set to 13×13. The accuracy with different patch size and different block size in the fruits boxes dataset is shown in Figure 10. It can be obtained that the accuracy tends to increase as the block size increases. Whereas larger block size will lose part of the features of the first-layer network [25], the block size is set to 5×5, and the block size is set to 8×8 in this paper.

4.2.3. Influence of Overlapping Ratio.
It has been verified that overlapping blocks not only improve target detection accuracy [50] but also resist geometric rotation and scaling changes to some extent. The robustness is also enhanced
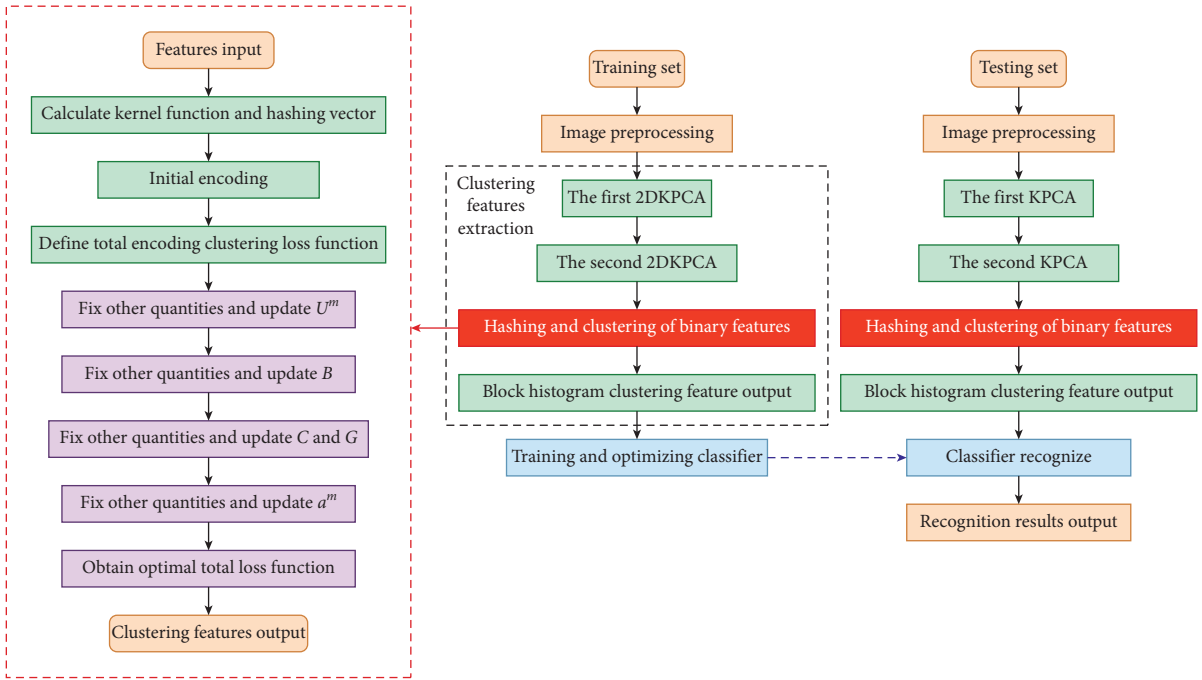
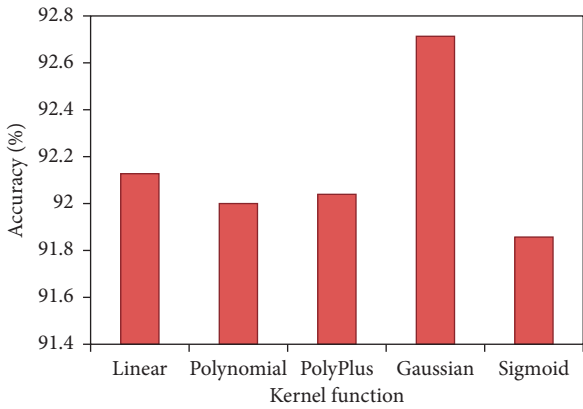FIGURE 7: Flowchart of fruits boxes identification.

FIGURE 8: Performance comparison of kernel functions.

FIGURE 9: Boxes recognition accuracy with different number of filters.

[51]. In order to strengthen the spatial information of fruits boxes, overlapping partitioning was carried out in this paper. The overlapping ratio of blocks was set from 0.1 to 0.9, respectively. The influence of overlapping ratio on fruits boxes recognition with the optimal other parameters was shown in Figure 11. It can be seen that when the block overlapping ratio is 0.6, the recognition performance of the model is optimal.

4.3. Analysis of Experimental Results. In order to verify the recognition ability of the proposed algorithm for fruit packing boxes, 80 images of each type of packing boxes were randomly selected as training samples, and the other 120 images were taken as test samples. The experiment was done with the optimal parameters. The recognition accuracies of different categories were shown in Table 1. The overall recognition meets the requirements of fruits boxes handling
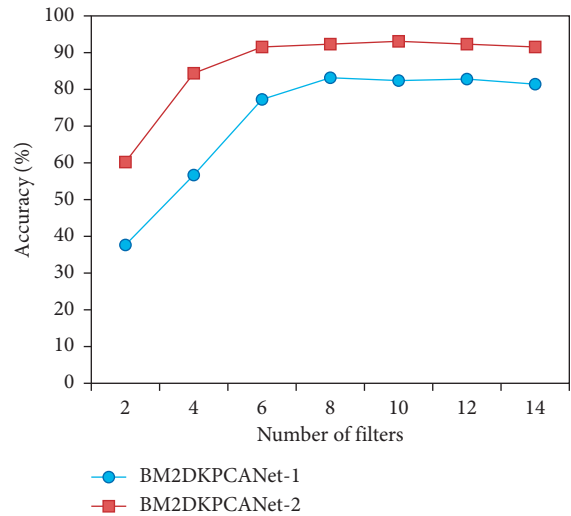
robot. For apple3, watermelon1, orange2, cantaloupe2, pear, and pineapple, etc., the top and side surfaces are easily to be confused, which lead to the lower accuracy. The average accuracy is 92.89%, which increased by 2.09% compared to the BMKPCANet [27] model.

This model was compared with PCANet [24], 2DPCANet [28], KPCANet [26], and BMKPCANet [27] in terms of recognition rate and time normalization to verify the performance of BM2DKPCANet model, as shown in Figure 12. It can be seen that although the proposed BM2DKPCANet model has more time consumption than the PCANet-related model, the recognition rate is 11.84% higher than the average of the PCANet-related model and
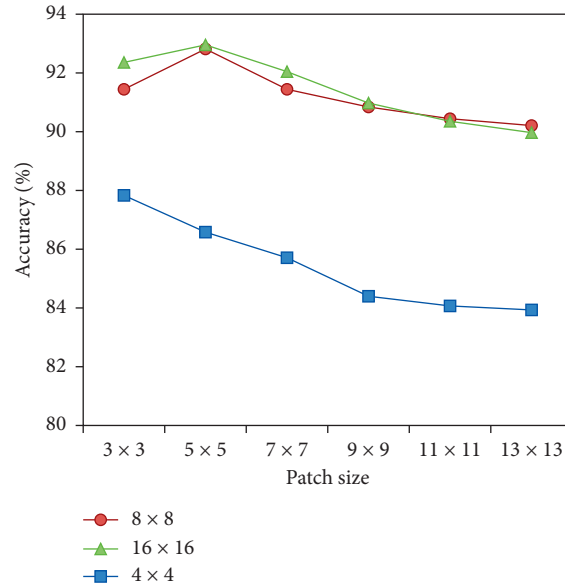
FIGURE 10: Influence of sampling block and histogram block on fruits boxes recognition.
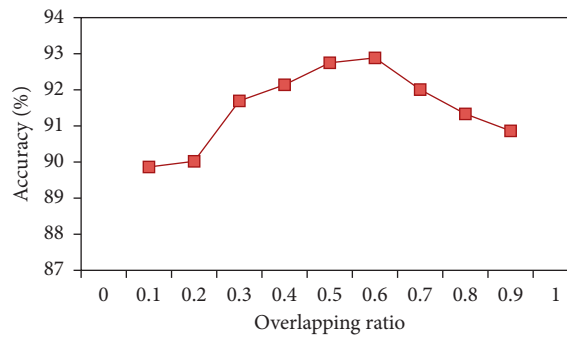

FIGURE 11: Influence of block overlap ratio on fruits boxes recognition.

TABLE 1: Recognition accuracy of different categories.

Categories	Right recognition number	Accuracy (%)
apple1	115	95.83
apple2	117	97.5
apple3	109	90.83
watermelon1	108	90
watermelon2	111	92.5
orange1	116	96.67
orange2	109	90.83
cantaloupe1	113	94.17
cantaloupe2	107	89.17
pomegranate	118	98.33
pear	106	88.33
durian	111	92.5
coconut	112	93.33
banana	111	92.5
pineapple	109	90.83

2.485% higher than the average of the KPCANet-related model. In addition, the time consumption of BM2DKPCANet can be saved by 24.5% on average. Taking that into account, BM2DKPCANET model is better than other models in fruits boxes recognition.

The recognition experiments were conducted with the same model parameters to verify the proposed multiview recognition algorithm. The comparisons of model performance on ETH-80 and COIL-100 are shown in Figures 13 and 14, respectively. It is easy to find that the proposed BM2DKPCANET model can achieve a higher recognition accuracy compared with other models.

It can be proved that the BM2DKPCANET model has achieved a good recognition effect for the three datasets in the part of the experimental results. Compared with PCANet and 2DPCANet models, the proposed model adopts kernel principal component analysis method, which makes the features to a high-dimensional space by nonlinear mapping and then carries out PCA dimensionality reduction. The nonlinear relationship of images is extracted, whereas the calculation is more complex and takes more time than PCA. The recognition accuracy is greatly improved. Compared with KPCA in KPCANet and BMKPCANet models, the proposed model does not need to transform two-
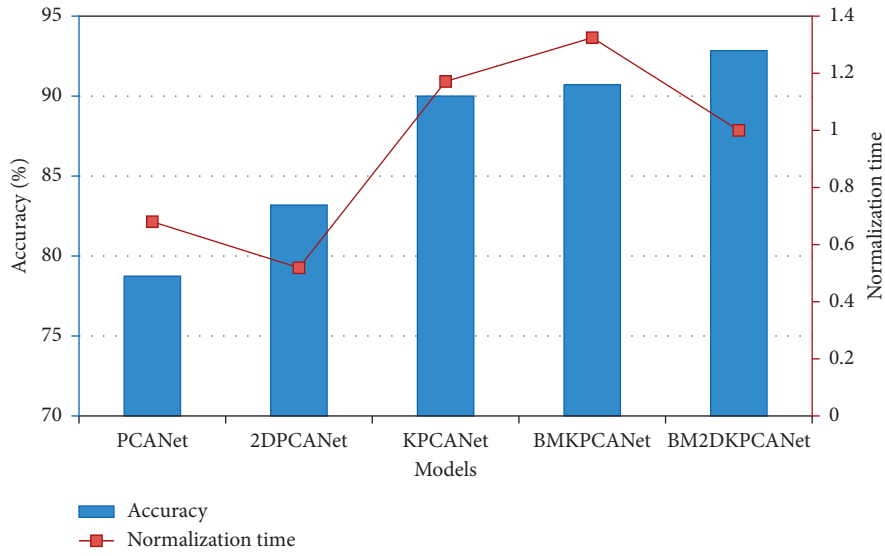
Figure 12: Comparison of model performance on the fruits boxes dataset.
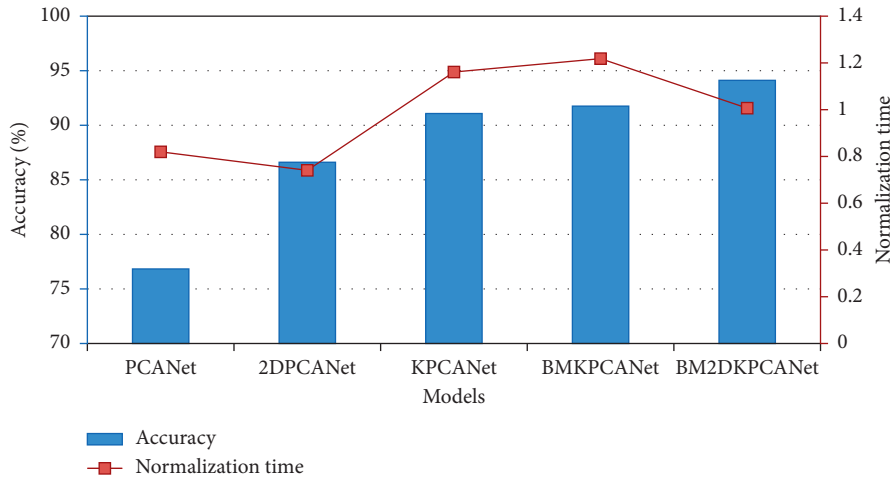

Figure 13: Comparison of model performance on the ETH-80 dataset.
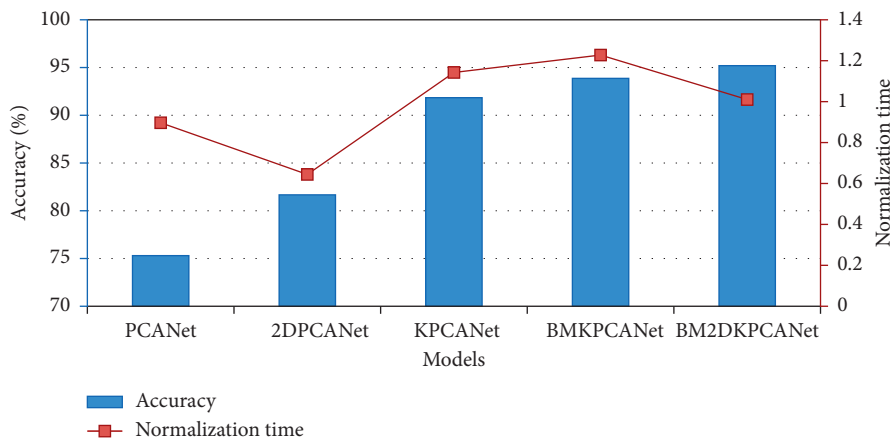

Figure 14: Comparison of model performance on the COIL-100 dataset.

dimensional matrix into one-dimensional vector but directly takes the average column vector method based on 2-dimensional image, which not only does not destroy the structural information of the original image as much as possible, but also greatly reduces the complexity. Therefore, not only is the recognition rate higher, but it also improves the efficiency.

5. Conclusions

In order to reduce the labour intensity of fruits handling in fruits orchards and fruits markets, this paper studied the fruits boxes recognition based on the machine vision. The recognition of 3D boxes was transformed into the feature extraction of 2D images. For the sake of the original 2D images' structures, the established BM2DKPCANet model performed two-layer 2DKPCA analysis on the 2D images. Binary clustering algorithm was added in the feature extraction stage to reduce the data redundancy caused by the multiview acquisition. The multiview recognition method for fruits boxes was proposed by combining BM2DKPCANet model with SVM classifier based on RBF. The experimental results showed that the recognition accuracy of this method is 11.84% higher than the average of PCANet model and 2.485% higher than the average of KPCANet model, which can meet the requirements of automatic rapid identification of fruits boxes handling. It laid a foundation for realizing the intelligent mechanization of fruits boxes handling and reducing the labour intensity of fruit farmers.

Data Availability

The dataset presented in this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

References

[1] H. Wang, R. Li, Y. Gao, C. Cao, L. Ge, and X. Xie, "Target recognition and localization of mobile robot with monocular PTZ camera," *Journal of Robotics*, vol. 2019, pp. 1–12, Article ID 8789725, 2019.

[2] J. Yuan, "Research progress analysis of robotics selective harvesting technologies," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 51, no. 9, pp. 1–17, 2020.

[3] H. Q. T. Ngo, V. N. S. Huynh, T. P. Nguyen, and H. Nguyen, "Sustainable agriculture: stable robust control in presence of uncertainties for multi-functional indoor transportation of farm products," *Agriculture*, vol. 10, no. 11, pp. 523–618, 2020.

[4] M. H. Li, *Object recognition and tracking based on structural sparse representation*, Ph.D. thesis, University of Electronic Science and Technology of China, Chengdu, China, 2020.

[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (ECCV)*, pp. 580–587, Columbus, OH, USA, June 2014.

[6] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, December 2015.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[8] M. Bucolo, L. Fortuna, M. Nelke, A. Rizzo, and T. Sciacca, "Prediction models for the corrosion phenomena in pulp & paper plant," *Control Engineering Practice*, vol. 10, no. 2, pp. 227–237, 2002.

[9] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 21–37, Amsterdam, Netherlands, October 2016.

[10] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, June 2016.

[11] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263–7271, Honolulu, HI, USA, July 2017.

[12] R. Joseph and F. Ali, "YOLOv3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767v1.

[13] A. Bochkovskiy, C. Y. Wang, and H. Y. Liao, "YOLOv4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.

[14] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multiview convolutional neural networks for 3d shape recognition," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 945–953, Washington, DC, USA, December 2015.

[15] Z. Gao, D. Wang, X. He, and H. Zhang, "Group-pair convolutional neural networks for multi-view based 3D object retrieval," in *Proceeding of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 1–8, New Orleans, LA, USA, February 2018.

[16] Z. Gao, D. Y. Wang, Y. B. Xue, G. P. Xu, H. Zhang, and Y. L. Wang, "3D object recognition based on pairwise multi-view convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 56, no. 10, pp. 305–315, 2018.

[17] Z. Gao, H. Xue, and S. Wan, "Multiple discrimination and pairwise CNN for view-based 3D object retrieval," *Neural Networks*, vol. 125, no. 2, pp. 290–302, 2020.

[18] H. Li, Y. Zheng, J. Cao, and Q. Cai, "Multi-view-based siamese convolutional neural network for 3D object retrieval," *Computers & Electrical Engineering*, vol. 78, no. 7, pp. 11–21, 2019.

[19] C. Lin and A. Kumar, "Contactless and partial 3D fingerprint recognition using multi-view deep representation," *Pattern Recognition*, vol. 83, pp. 314–327, 2018.

[20] Y. Yang, F. Chen, F. Wu, D. Zeng, Y.-M. Ji, and X.-Y. Jing, "Multi-view semantic learning network for point cloud based

3D object detection," *Neurocomputing*, vol. 397, no. 3, pp. 477–485, 2020.

[21] Y. Zhang, X. S. Guo, X. Guo, H. Ren, and L. Li, "Multi-view classification with semi-supervised learning for SAR target recognition," *Signal Processing*, vol. 183, Article ID 108030, 2021.

[22] X. Chen, Y. Chen, and H. Najjaran, "End-to-end 3D object model retrieval by projecting the point cloud onto a unique discriminating 2D view," *Neurocomputing*, vol. 402, pp. 336–345, 2020.

[23] M. Bucolo, A. Buscarino, C. Famoso, L. Fortuna, and M. Frasca, "Control of imperfect dynamical systems," *Nonlinear Dynamics*, vol. 98, no. 4, pp. 2989–2999, 2019.

[24] L. Fei, B. Zhang, J. Wen, S. Teng, S. Li, and D. Zhang, "Jointly learning compact multi-view hash codes for few-shot FKP recognition," *Pattern Recognition*, vol. 115, Article ID 107894, 2021.

[25] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: a simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[26] D. Wu, J. S. Wu, R. Zeng, L. Y. Jiang, L. Senhadji, and H. Z. Shu, "Kernel principal component analysis network for image classification," *Journal of Southeast University(English Edition)*, vol. 31, no. 4, pp. 469–473, 2015.

[27] X. N. Li, H. Wu, and X. H. Yang, "Multi-view recognition of fruit packing boxes based on features clustering angle," *High Technology Letters*, vol. 47, 2021.

[28] D. Yu and X.-J. Wu, "2DPCANet: a deep leaning network for face recognition," *Multimedia Tools and Applications*, vol. 77, no. 10, pp. 12919–12934, 2018.

[29] D. Zhang, S. Chen, and Z. Zhou, "Recognizing face or object from a single image: linear vs. kernel methods on patterns," in *Proceedings of the Joint IAPR International Workshops on Structural and Syntactic Pattern Recognition and Statistical Techniques in Pattern Recognition, LNCS 4109*, pp. 889–897, Hong Kong, China, August 2006.

[30] J. Jian Yang, D. Zhang, A. F. Frangi, and J. Y. Jing-yu Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[31] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[32] V. D. M. Nhat and S. Lee, "Kernel-based 2DPCA for face recognition," in *Proceedings of the 2007 IEEE International Symposium on Signal Processing and Information Technology*, pp. 35–39, Giza, Egypt, December 2007.

[33] S. Zhou, Y. Zheng, and X. Mu, "K2DPCA methods for face recognition based on Cholesky decomposition," *Systems Engineering-Theory&Practice*, vol. 36, no. 2, pp. 528–535, 2016.

[34] M. Xu, D. Xu, and M. Wei, "Face recognition with Laplacian eigenmaps based on 2D-KPCA," *Application Research of Computers*, vol. 34, no. 7, pp. 2212–2215, 2017.

[35] Y. Choi, S. Ozawa, and M. Lee, "Incremental two-dimensional kernel principal component analysis," *Neurocomputing*, vol. 134, pp. 280–288, 2014.

[36] M. Esmaeili, M. Ahmadi, and A. Kazemi, "Kernel-based two-dimensional principal component analysis applied for parameterization in history matching," *Journal of Petroleum Science and Engineering*, vol. 191, Article ID 107134, 2020.

[37] Y. Xiang, G. Yang, J. Zhang, and Q. Wang, "Dimensionality reduction for hyperspectral image using a segmented row-column kernel two-dimensional principal component analysis method," *Infrared Technology*, vol. 39, no. 12, pp. 1107–1113, 2017.

[38] N. Sun, H.-x. Wang, Z.-h. Ji, C.-r. Zou, and L. Zhao, "An efficient algorithm for kernel two-dimensional principal component analysis," *Neural Computing and Applications*, vol. 17, no. 1, pp. 59–64, 2007.

[39] A. Eftekhari, M. Forouzanfar, H. Abrishami Moghaddam, and J. Alirezaie, "Block-wise 2D kernel PCA/LDA for face recognition," *Information Processing Letters*, vol. 110, no. 17, pp. 761–766, 2010.

[40] L. Wang and X. Zhou, "Approximation kernel 2DPCA by mixture of vector and matrix representation," *International Conference on Computational Intelligence and Security*, vol. 12, no. 3, pp. 1298–1302, 2011.

[41] B. Chen, J. Yang, B. Jeon, and X. Zhang, "Kernel quaternion principal component analysis and its application in RGB-D object recognition," *Neurocomputing*, vol. 266, pp. 293–303, 2017.

[42] B. J. Chen, J. H. Yang, C. N. Fan, Q. T. Su, and D. C. Wang, "Block-wise two dimensional kernel quaternion principal component analysis," *Journal of Beijing University of Posts and Telecommunications*, vol. 42, no. 1, pp. 53–60, 2019.

[43] H. Peng, S. K. Huang, L. Tao, and T. X. Zhang, "Multi-view modeling of 3D target based on Zernike moments," *Infrared and Laser Engineering*, vol. 34, no. 3, pp. 292–296, 2005.

[44] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 409–415, Madison, WI, USA, June 2003.

[45] S. A. Nene, S. K. Nayar, and H. Murase, *Columbia Object Image Library (COIL 20), CUCS-005-96*, New York: Department of Computer Science, New York, NY, USA, 1996.

[46] J. R. Hu and Z. Z. Yu, "Model validation method with multivariate output based on kernel principal component analysis," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 43, no. 7, pp. 1470–1480, 2017.

[47] Z. Zhang, L. Liu, F. Shen, H. T. Shen, and L. Shao, "Binary multi-view clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1774–1782, 2019.

[48] X. Zhang, Z. Yang, and S. Cao, "Spectral detection method for chilling damage of sweet potato based on support vector machine," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 51, no. 2, pp. 471–477, 2020.

[49] H. Zhang, H. Sun, and P. Shi, "Chip appearance inspection method for high-precision SMT equipment," *Machines*, vol. 9, no. 2, p. 34, 2021.

[50] R. H. Miao, H. Yang, J. L. Wu, and H. Y. Liu, "Weed identification of overlapping spinach leaves based on image sub-block and reconstruction," *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, vol. 36, no. 4, pp. 178–184, 2020.

[51] Y. Wang, X. Kang, and Y. Chen, "Robust and accurate detection of image copy-move forgery using PCET-SVD and histogram of block similarity measures," *Journal of Information Security and Applications*, vol. 54, no. 10, Article ID 102536, 2020.