Research on Feature Fusion Technology of Fruit and Vegetable Image Recognition Based on SVM

Yanqing Wang[1(✉)], Yipu Wang[1,2], Chaoxia Shi[3], and Hui Shi[4]

[1] School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China
wyq0325@126.com

[2] ZhongJing FuDian (Shanghai) Electronic Technology Co., Ltd., Shanghai 201203, China

[3] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

[4] Henan Electric Power Transmission and Transformation Engineering Company, Zhengzhou 450051, China

Abstract. In order to improve the accuracy and stability of fruit and vegetable image recognition by single feature, this project proposed multi-feature fusion algorithms and SVM classification algorithms. This project not only introduces the Reproducing Kernel Hilbert space to improve the multi-feature compatibility and improve multi-feature fusion algorithm, but also introduces TPS transformation model in SVM classifier to improve the classification accuracy, real-time and robustness of integration feature. By using multi-feature fusion algorithms and SVM classification algorithms, experimental results show that we can recognize the common fruit and vegetable images efficiently and accurately.

Keywords: Feature extraction · Multi-feature fusion · Support vector machine · Fruit and vegetable image recognition

1 Introduction

With the development of technology in the field of pattern recognition, computer vision and image processing has become the hot point of research. Because the technology can identify various types of images automatically and get result precisely and efficiently. So domestic and foreign researchers conduct to identify fruit and vegetable mostly based on contour feature, color feature, texture feature and so on. When researchers identify fruit and vegetable images in uncontrolled lighting conditions and the context of complex environments, it will be lead to increase the likelihood of misclassification. Although the multi-feature algorithms achieved some results, but most of multi-feature algorithms just combine some features simply without effective integration. So, the number of feature vectors' dimension is too large and the accuracy of image recognition is poor. Currently, there are some kinds of shortcomings in fruit and vegetable recognition filed. Some algorithms can find the position [1, 2] or grade [3] of fruit or vegetable, but these algorithms can't identify the type of fruit and vegetable. Some algorithms can identify

W. Che et al. (Eds.): ICYCSEE 2016, Part I, CCIS 623, pp. 591–599, 2016.
DOI: 10.1007/978-981-10-2053-7_52

the type of fruit and vegetable by only a single feature vector [4–6], so the recognition accuracy of these algorithms is low. This paper proposes an improved heterogeneous features compatibility feature fusion algorithm and an improved SVM algorithm to achieve multi-classification of fruits and vegetables.

2 Feature Extraction and Feature Fusion

This paper used mature feature extraction algorithms to extract features. And then, we combined multiple features with one feature. The feature fusion algorithm can improve the performance of the system. An important issue is the integration of features compatibility between heterogeneous features. Therefore, we need use normalization techniques to solve the problems [7] before connecting vector. This project introduces the Reproducing Kernel Hilbert space to improve the multi-feature compatibility and improve the accuracy of multi-feature fusion algorithm.

2.1 Feature Extraction

This paper mainly extracted contour feature, color feature and texture feature. OTSU algorithm [8, 9] can segment original image into foreground image and background image by a threshold. The OTSU threshold can be used as the input threshold of Canny algorithm [10, 11], and then we can get contour extraction. This paper used K-means algorithm [12, 13] to extract color range of fruit and vegetable images. We can obtain the color features by the cluster of similar color. This paper used roughness component, contrast components, and orientation components of Tamura texture feature [14] as texture features of images.

2.2 Feature Fusion

Change feature compatibility. This paper introduces the Reproducing Kernel Hilbert space to improve the multi-feature compatibility and improve multi-feature fusion algorithm, and then we can fuse features. We can get two sets of heterogeneous features $X = \left\{x_i \in \mathbb{R}^{D_x},\ i = 1, \ldots, N\right\}$, $Y = \left\{y_i \in \mathbb{R}^{D_y},\ i = 1, \ldots, N\right\}$ and constraints set S. If $x_i \in X$ and $y_j \in Y$ have the same kind of labels, then $(i,j) \in S$. We defined feature spaces $\mathcal{F}_1$ and $\mathcal{F}_2$ with nonlinear mapping $\phi, \mathbb{R}^{D_x} \to \mathcal{F}_1, \mathbb{R}^{D_y} \to \mathcal{F}_2$. We introduces the Reproducing Kernel Hilbert space $K(\cdot,\cdot)$. $K(\cdot,\cdot)$ is a semi-definite kernel function, and then $\left\langle \phi(x_i),\ \phi\left(x_j\right)\right\rangle = K\left(x_i,\ x_j\right)$ and $\left\langle \phi\left(x_j\right),\ \phi(x_i)\right\rangle = K\left(y_i,\ y_j\right)$ are established. The purpose of introduce Reproducing Kernel Hilbert space is to find a pair of mapping $\mathrm{P_x}$ and $\mathrm{P_y}$, then heterogeneous data sets are the closest set S after mapping to get the optimal solution of formula (1).j

$$\underset{P_x,P_y}{\arg\min}\, J\left(P_x,\ P_y\right) = \sum_i \sum_j \left\| P_x^T \phi(x_i) - P_y^T \phi\left(y_j\right)\right\|^2 S_{ij} \tag{1}$$

Formula (1) can be obtained the formula (2) after rewriting and simplify.

$$J(P_x, P_y) = Tr\left\{ \begin{bmatrix} P_x \\ P_y \end{bmatrix}^T \begin{bmatrix} \phi(X) & \\ & \phi(Y) \end{bmatrix} \left(D - \begin{bmatrix} 0 & S \\ S^T & 0 \end{bmatrix} \right) \begin{bmatrix} \phi(X) & \\ & \phi(Y) \end{bmatrix}^T \begin{bmatrix} P_x \\ P_y \end{bmatrix} \right\} \quad (2)$$

In formula (2), $\Phi(X) = [\phi(x_1),\ \phi(x_2),\ \ldots,\ \phi(x_N)]$, $\Phi(Y) = [\phi(y_1),\ \phi(y_2),\ \ldots,\ \phi(y_N)]$. We defined that $W = \begin{bmatrix} 0 & S \\ S^T & 0 \end{bmatrix}$, then D is a diagonal matrix which values in each row equals the sum value of the row of W. We defined $P_x = \left[p_x^1,\ p_x^2, \ldots,\ p_x^m\right]$ and $P_y = \left[p_y^1,\ p_y^2, \ldots,\ p_y^m\right]$, p_x^i and p_y^j can be extended as $p_x^i = \sum_{j=1}^N \alpha_j^i \phi(x_i)$ and $p_y^i = \sum_{j=1}^N \beta_j^i \phi(y_i)$ linearly according to theorem. We defined that $\alpha^i = \left[\alpha_1^i,\ \alpha_2^i, \ldots,\ \alpha_N^i\right]^T \in \mathbb{R}^{N\times 1}$, $\beta^i = \left[\beta_1^i,\ \beta_2^i, \ldots, \beta_N^i\right]^T \in \mathbb{R}^{N\times 1}$, $A = [\alpha^1,\ \alpha^2, \ldots, \alpha^m] \in \mathbb{R}^{N\times m}$ and $B = \left[\beta^1, \beta^2, \ldots, \beta^m\right] \in \mathbb{R}^{N\times m}$, then we can get formulas (3) and (4)

$$P_x = \Phi(X)A \quad (3)$$

$$P_y = \Phi(Y)B \quad (4)$$

We can get formula (5) by formulas (2), (3) and (4).

$$J(A, B) = Tr\left\{ \begin{bmatrix} A \\ B \end{bmatrix}^T \begin{bmatrix} K_x & \\ & K_y \end{bmatrix} \left(D - \begin{bmatrix} 0 & S \\ S^T & 0 \end{bmatrix} \right) \begin{bmatrix} K_x & \\ & K_y \end{bmatrix}^T \begin{bmatrix} A \\ B \end{bmatrix} \right\} \quad (5)$$

K_x and K_y are nuclear matrixes, $K_x(i,j) = \mathcal{K}(x_i, x_j)$, $K_y(i,j) = \mathcal{K}(y_i, y_j)$. We defined that $P = \begin{bmatrix} A \\ B \end{bmatrix}$ and $K = \begin{bmatrix} K_x & \\ & K_y \end{bmatrix}$, and then we can get formula (6).

$$J(P) = Tr\left\{P^T K(D - W)K^T P\right\} \quad (6)$$

Finally, in order to remove any scaling factor we add a constraint $P^T KDK^T P = I$. Then we can get formula (7).

$$P^* = \underset{P^T KDK^T P = I}{\arg\min}\ Tr\left\{P^T K(D - W)K^T P\right\} = \underset{P^T KDK^T P = I}{\arg\max}\ Tr\left\{P^T KWK^T P\right\} \quad (7)$$

By solving the eigenvalue of maximum problem of formula (8), we can get the optimal solution of P.

$$\left(KWK^T\right)P = \left(KDK^T\right)P\Lambda \quad (8)$$

Feature Fusion. We defined two sets of heterogeneous features are $X \in \mathbb{R}^{D_x \times N}$ and $X \in \mathbb{R}^{D_x \times N}$, then we can get features $X^p \in \mathbb{R}^{m\times N}$ and $Y^p \in \mathbb{R}^{m\times N}$ through mapping by formulas (9) and (10).

$$X^p = P_x^T \Phi(X) = A^T \Phi(X)^T \Phi(X) = A^T K_x \tag{9}$$

$$Y^p = P_y^T \Phi(Y) = B^T \Phi(Y)^T \Phi(Y) = B^T K_y \tag{10}$$

We can simply get the fusion feature $Z \in \mathbb{R}^{m \times N}$ of train set through calculating the average feature points of the training set as formula (11).

$$Z = \frac{1}{2}(X^p + Y^p) \tag{11}$$

We defined heterogeneous feature set of test $X_t = \left\{x_i^t \in \mathbb{R}^{D_x}, i = 1, \ldots, N_t\right\}$ and $Y_t = \left\{y_i^t \in \mathbb{R}^{D_y}, i = 1, \ldots, N_t\right\}$. The test feature also can mapping into the same space according to formulas (12) and (13).

$$X_t^P = P_x^T \Phi(X_t) = A^T \Phi(X)^T \Phi(X_t) = A^T K_x^t \tag{12}$$

$$Y_t^P = P_y^T \Phi(Y_t) = B^T \Phi(Y)^T \Phi(Y_t) = B^T K_y^t \tag{13}$$

K_x^t and K_y^t are nuclear matrixes, $K_x^t(i,j) = \mathcal{K}\left(x_i, x_j^t\right)$, $K_y^t(i,j) = \mathcal{K}\left(y_i, y_j^t\right)$. We can calculate fusion features by the formula (14).

$$Z_t = \frac{1}{2}\left(X_t^p + Y_t^p\right) \tag{14}$$

Finally, this paper used SVM classification $\hat{f}$ to train fusion train set Z and label. For test features X_t and Y_t, we can calculate the fusion feature Z_t and predict the classification label $\hat{f}(Z_t)$.

3 Improved SVM Algorithm

This paper used SVM classifier to classify. Since we used multi-feature fusion, we should improve SVM algorithm and improve the accuracy of results. This paper introduces the thin-plate splines (TPS) conversion model in SVM classifier, which has a strong versatility and power performance in terms of higher order distortion.

We defined the mapping function of TPS conversion model is $f(x) : \mathbb{R}^d \to \mathbb{R}^d$. So the minimizing the penalty function for the smooth TPS is formula (15).

$$J_m^d(f) = \int \|D^m f\|^2 dX = \sum_{\alpha_1 + \ldots + \alpha_d = m} \frac{m!}{\alpha_1! \ldots \alpha_d!} \int \cdots \int \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}\right)^2 \prod_{j=1}^{d} dx_j \tag{15}$$

$D^m f$ is the m order partial derivatives of matrix f. α_k is a positive number. $dX = \prod_{j=1}^{d} dx_j$, x_j is a part of vector x. The classic solution of formula (15) has radial basis function form (TPS interpolation function) as shown in formula (16).

$$f_k(x) = \sum_{i=1}^{n} \psi_i G(\|\mathbf{x} - \mathbf{x}_i\|) + \ell^T \mathbf{x} + c \tag{16}$$

$\|\cdot\|$ is the Euclidean norm. $\{\psi_i\}$ is the weight of nonlinear part. ℓ and c are the weight of linear part. We can simplify formula (15) by the radial distance of TPS kernel, as shown in formula (17).

$$G(\mathbf{x}, \mathbf{x}_i) = G(\|\mathbf{x} - \mathbf{x}_i\|) \propto \begin{cases} \|\mathbf{x} - \mathbf{x}_i\|^{2m-d} \ln\|\mathbf{x} - \mathbf{x}_i\| & 2m - diseven \\ \|\mathbf{x} - \mathbf{x}_i\|^{2m-d} & others \end{cases} \tag{17}$$

The input space of interpolation point TPS transformation model (as shown in formula (16)) is the deformation of nonlinear learning distortion. This shift should be able to ensure the desired smoothness. Accordingly, as shown in formula (15) we reduced the bending energy $J_m^d(f)$ on maximum degree. In the settings of learning measure, x is one of train fusion feature. x is converted to $f(x)$. By calculation, the matrix can be obtained in the form of formula (18).

$$f(\mathbf{x}) = L\mathbf{x} + \psi \begin{pmatrix} G(\mathbf{x}, \mathbf{x}_1) \\ \cdots \\ G(\mathbf{x}, \mathbf{x}_p) \end{pmatrix} = L\mathbf{x} + \psi \mathbf{G}(\mathbf{x}) \tag{18}$$

L is the d*d linear transformation matrix, ψ is the d*p nonlinear weight matrix, p is the number of anchors. By nonlinear TPS transformation model, SVM paradigm can be defined by the Margin-Radius-Ratio (MRR).

We defined train set $\chi = \{\mathbf{x}_i | \mathbf{x}_i \in \mathbb{R}^d, i = 1, \ldots, n\}$ and classification label y_i. We can get formula (19) by formula (18).

$$\begin{aligned} & \min_{L,\psi,w,b} J = \frac{1}{2}\|\mathbf{w}\|^2 + C_1 \sum_{i=1}^{n} \xi_i + C_2 \|\psi\|_F^2 \\ s.t.\ & y_i(\mathbf{w}^T f(\mathbf{x}_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i = 1\ldots n(I \,\&\, II) \\ & \|f(\mathbf{x}_i) - \mathbf{x}_c\|^2 \leq 1, \ \ \forall i = 1\ldots n(III) \\ & \sum_{i=1}^{P} \psi_i^k = 0, \sum_{i=1}^{P} \psi_i^k \mathbf{x}_i^k = 0, \ \forall k = 1\ldots d(IV) \end{aligned} \tag{19}$$

ψ_i^k is k-th column of ψ.$\mathbf{x}^k$ is k-th part of $\mathbf{x}$. In addition to traditional soft margin of SVM components, another component is $\|\psi\|_F^2$. Joining squared Frobenius norm ψ in the objective function can be regularized objective function to prevent overfitting. C_1 and C_2 are two weigh hyperparameters. The first two non-constraint (I and II) are the same with traditional SVM same. The third constraint is a non-closed unit ball constrained, so its transformation space in a unit radius sphere can avoid the trivial solution, and it is the center of all samples. The last two equivalent constraint (IV) can be used to maintain TPS transform properties at infinity. This paper introduces TPS model and improve the classification accuracy of the results.

4 Experimental Results

In classification step, we chose 11 kinds of fruits and vegetables image (apple, banana, carrot, cucumber, kiwi fruit, orange, yellowish orange, pear, eggplant, tomato, date). The number of training samples and test samples image is shown in Table 1.

Table 1. The number of training samples and test samples image

Recognition type	Training sample	Test sample
Apple	96	48
Banana	98	48
Carrot	78	39
Cucumber	67	33
Kiwi fruit	53	26
Orange	118	58
Yellowish orange	104	52
Pear	110	55
Eggplant	87	43
Tomato	51	25
Date	46	23
Total	908	450

We can get contour feature, color feature and texture feature. We used traditional SVM algorithm to recognition by single features and fusion features respectively. We used traditional SVM algorithm and improved SVM algorithm respectively to recognition by fusion features. During the experiment, we selected polynomial kernel as the SVM kernel function, and used PSO algorithm to determine punishment variables c and gamma function g: c = 54.3, g = 64.

The result of classification algorithms is shown in Table 2.

We can know some information through Table 2.

1. The recognition rate of the same test sample are very different between different single feature. Figure 1 is the contour feature comparison figure of test sample. Figure 1 (a) and (c) are the carrot and the cucumber test image respectively. Figure 1(b) and (d) are the carrot and the cucumber contour image respectively. We can find the contour feature of the two kinds vegetables are very similar from the figure, so it's so easy to confuse by contour feature. Figure 2 is the color feature comparison figure of test sample. Both of Fig. 2(a) and (c) are apple test images. Figure 2(b) is the pear test image. Figure 2(d) is the tomato test image. We can find that different kinds of apples have great color differences, and then the difference increase the difficulty of classification.

2. As 1. above, some features have high similarity and easily confused. There are many kinds of similar images, so the classification is very difficult. An improved SVM classification algorithm in this paper can distinguish similar images, and the classification accuracy rates are over 95 %.

Table 2. The result of classification algorithms

Recognition type	Recognition rate of single feature			Recognition rate of fusion feature	
	Contour	Color	Texture	Traditional SVM	Improved SVM
Apple	93.75 %	91.66 %	89.58 %	93.75 %	95.83 %
Banana	100 %	95.83 %	91.66 %	100 %	100 %
Carrot	94.87 %	97.43 %	92.30 %	100 %	100 %
Cucumber	90.90 %	93.93 %	90.90 %	93.93 %	96.96 %
Kiwi fruit	88.46 %	92.30 %	84.61 %	92.30 %	96.15 %
Orange	91.37 %	91.37 %	89.65 %	91.37 %	94.82 %
Yellowish orange	92.30 %	94.23 %	90.38 %	94.23 %	96.15 %
Pear	96.36 %	98.18 %	92.72 %	98.18 %	100 %
Eggplant	95.34 %	97.67 %	93.02 %	97.67 %	100 %
Tomato	92 %	92 %	88 %	92 %	96 %
Date	86.95 %	91.30 %	86.95 %	91.30 %	95.65 %
Average	92.93 %	94.17 %	89.97 %	94.97 %	97.41 %

Fig. 1. The contour feature comparison of test sample

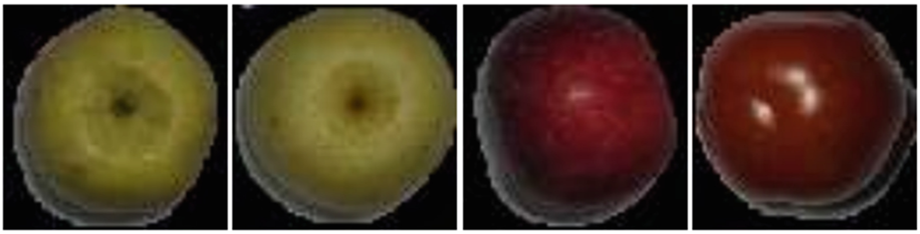

Fig. 2. The color feature comparison of test sample

3. In case of using the same kind of SVM algorithm, the classification accuracy rates with fusion feature are higher than the classification accuracy rates with single feature. The result indicates that the fusion feature utilize more information and enhance the ability to identify.

4. Improved SVM algorithm added TPS transformation model, since this model has a stronger versatility and power performance in higher-order deformation terms, as compared with the traditional SVM algorithm, classification accuracy increased by nearly 4 %.

Thus, the following conclusions:

1. The accuracy rate of single feature recognition is lower, and the reliability and stability of classification is poor. The accuracy rate of fusion feature recognition is higher than simple feature recognition and it has good robustness.
2. Multi-feature fusion algorithm can obtain more information. The fusion feature recognition rate is greatly improved compared with the single feature recognition rate.
3. The classification accuracy of improved SVM algorithm with fusion features is improved greatly, and it can compare very similar images. The average recognition rate with fusion features is 97.41 %.

5 Conclusion

This project not only introduces the Reproducing Kernel Hilbert space to improve the multi-feature compatibility and improve multi-feature fusion algorithm, but also introduces TPS transformation model in SVM classifier to improve the classification accuracy, real-time and robustness of integration feature. The average recognition rate with fusion features is 97.41 %.

Acknowledgments. This paper has been supported by the National Natural Science Foundation of China (Grant No. 61371040).

References

1. Gatica, C., Best, S., Ceroni, J., Lefranc, G.: A new method for olive fruits recognition. In: San Martin, C., Kim, S.-W. (eds.) CIARP 2011. LNCS, vol. 7042, pp. 646–653. Springer, Heidelberg (2011)
2. Amosh, L.S., Sheikh Abdullah, S.N.H., Che Mohd, C.R., Jameson, J.: Adaptive region growing for automated oil palm fruit quality recognition. In: Zaman, H.B., Robinson, P., Olivier, P., Shih, T.K., Velastin, S. (eds.) IVIC 2013. LNCS, vol. 8237, pp. 184–192. Springer, Heidelberg (2013)
3. Fang, C., Hua, C.: Advances in the application of image processing fruit grading. In: Li, D., Chen, Y. (eds.) CCTA 2013. IFIP AICT, vol. 419, pp. 168–174. Springer, Heidelberg (2013)
4. Zawbaa, H.M., Abbass, M., Hazman, M., Hassenian, A.E.: Automatic fruit image recognition system based on shape and color features. In: Hassanien, A.E., Tolba, M.F., Taher Azar, A. (eds.) AMLTA 2014. CCIS, vol. 488, pp. 278–290. Springer, Heidelberg (2014)

5. Potter, M.C., Hagmann, C.E.: Banana or fruit? Detection and recognition across categorical levels in RSVP. J. Psychon. Bull. Rev. **22**, 578–585 (2015)
6. García-Lamont, F., Cervantes, J., Ruiz, S., López-Chau, A.: Color characterization comparison for machine vision-based fruit recognition. In: Huang, D.-S., Bevilacqua, V., Premaratne, P. (eds.) ICIC 2015. LNCS, vol. 9225, pp. 258–270. Springer, Heidelberg (2015)
7. Jain, A., Nandakuma, K., Ross, A.: A. Score normalization in multimodal biometric systems. J Pattern Recogn. **38**, 2270–2285 (2005)
8. Cheng, C.-C., Tsai, C.-M.: Using red-otsu thresholding to detect the bus routes number for helping blinds to take bus. In: Ali, M., Pan, J.-S., Chen, S.-M., Horng, M.-F. (eds.) IEA/AIE 2014, Part I. LNCS, vol. 8481, pp. 321–330. Springer, Heidelberg (2014)
9. Martín-Rodríguez, F.: New tools for gray level histogram analysis, applications in segmentation. In: Kamel, M., Campilho, A. (eds.) ICIAR 2013. LNCS, vol. 7950, pp. 326–335. Springer, Heidelberg (2013)
10. Ramachandran, D., Kam, M., Chiu, J.: Social dynamics of early stage co-design in developing regions. J. Proc. Chi 1087–1096 (2015)
11. Shah, Z.H., Kaushik, V.: Performance analysis of canny edge detection for illumination invariant facial expression recognition. In: IEEE International Conference on Industrial Instrumentation and Control (ICIC), pp. 101–106 (2015)
12. Kaur, R., Heena, H.: An approach defining gait recognition system using k-means and MDA. Int. J. Comput. Appl. **117**, 1–4 (2015)
13. Zhao, C., Li, X., Yan, C.: Bisecting k-means clustering based face recognition using block-based bag of words model. J Optik – Int. J. Light Electron Optics **126**, 1761–1766 (2015)
14. Tiwari, D., Tyagi, V.: Dynamic texture recognition based on completed volume local binary pattern. J. Multidimension. Syst. Signal Process. **27**, 1–13 (2015)