IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 24, 2019, accepted October 30, 2019, date of publication November 4, 2019,
date of current version November 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951473

Ellipsoid Fitting Using Variable Sample Consensus and Two-Ellipsoid-Bounding-Counting for Locating Lingwu Long Jujubes in a Natural Environment

MIN HAN[1], JIANGMING KAN[1], AND YUTAN WANG[2]
[1]College of Engineering, Beijing Forestry University, Beijing 100083, China
[2]College of Mechanical Engineering, Ningxia University, Yinchuan 750000, China

Corresponding author: Jiangming Kan (kanjm@bjfu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 31660239.

ABSTRACT Locating Lingwu Long Jujubes is a key step for the automatic picking of the jujubes which can lower labour costs. In this paper, a method for detecting Lingwu Long Jujubes in a natural environment with a three-dimensional (3D) point cloud is proposed. First, the jujubes are preliminarily extracted in two-dimensional (2D) image. Then, the points are fitted to ellipsoid by least square (LS) in the sample consensus framework. The model scores are different when sample size is changed. The variable sample consensus (VARSAC) is proposed to get higher score than random sample consensus (RANSAC). The normal vector and the distance need to be calculated in score calculation of the RANSAC. However, this score calculation method is complex and time-consuming. Thus, a new method called two-ellipsoid-bounding-counting (TEBC) is proposed. The TEBC produces two auxiliary ellipsoids that are obtained by scaling semiaxis of the model. The points, which is bounded by two auxiliary ellipsoids, are regarded as inliers. The functional value of every candidate point is calculated to select the inliers. The valid ellipsoids are determined by the prior information and the invariants. Finally, the centre, size and the attitude angle of the jujubes are solved using eigenvalues and eigenvectors. Experiments are carried out on synthetic and real datasets. The experimental results show that the proposed method can faster and more accurately detect jujube. The speed of the VARSAC+TEBC is approximately 4 times faster than that of the RANSAC in the real dataset.

INDEX TERMS Ellipsoid fitting, object locating, sample size, two-ellipsoid-bounding-counting.

I. INTRODUCTION

Lingwu Long Jujubes are important economical fruits in Ningxia, China. To improve picking efficiency and reduce labour costs, there is an increasing demand for automatic picking techniques. Locating the fruits in a natural environment is a key step for automatic picking. The premise of movement of manipulator is to determine the position of Lingwu Long Jujubes. In addition, the attitude angle and size of the jujube are also useful information for efficient picking. Target detection based on vision is very promising. Significant achievements have been realized using two-dimensional (2D) image and three-dimensional (3D) point cloud for fruit detection. In [1], two levels of visual characterizations are used for extraction according to information of depth and color images. This method is successful at recognizing green apples in a canopy foreground. Jia *et al.* [2] use the K-means cluster segmentation and optimized radial basis function (RBF) neural networks to improve the recognition accuracy and speed of an apple picking robot. Reference [3] presented a robotic arm that was controlled by a depth sensor. The robotic arm can be used to pluck fruit or prune branches using a proper microcontroller board. The detection accuracy of a harvesting robot is guaranteed due to the development of vision hardware. Binocular stereo vision can obtain 3D points clouds using depth maps, but its accuracy depends on stereo matching. A Kinect combines an infrared sensor and an RGB camera. The RGB camera is used to obtain 2D image

The associate editor coordinating the review of this manuscript and approving it for publication was Tallha Akram.

and the depth sensor is used to obtain the 3D coordinates of points. The colour of a 2D image can be mapped to the 3D points using their mapping relation. The 3D points contain the deep information of the objects in the real world. Thus, the Kinect has more advantages than a general CCD camera when applied in harvesting robots.

Statistical analysis shows that the shape of Lingwu Long Jujubes is similar to an ellipsoid. Locating Lingwu Long Jujubes for harvest robots is a problem of the fitting of ellipsoidal objects in a natural environment. The fitting of some geometric primitives, such as planes, spheres, cylinders, and cones, is simple and well-developed [4]. However, the fitting of ellipsoid is slow and difficult because of the greater number of parameters. In fact, an ellipsoid with three semiaxis parameters a, b, and c is a better model than a sphere with only one radius parameter r or a cylinder, which can be regarded as special case of an ellipsoid when one semiaxis is infinity. Therefore, an ellipsoid is the best of the geometric primitives to fit objects in the real world. However, the fitting of an ellipsoid with rotation and translation is complex and sometimes ambiguous [5]. The specific fitting methods of ellipsoid include the least square (LS) [6], the sum of discriminants method (SOD) [7], the ellipsoid-specific method (HES) [8] and a series of variations that set different quadratic constraints. This paper adopts the LS for the fitting of the ellipsoid because this method is easy to implement.

The fitting method of geometric primitives can be divided into direct fitting and indirect fitting. Direct fitting directly fits all points, while indirect fitting often adopts the sample consensus framework [9]. The RANSAC is a very classic framework of sample consensus and there are many derived algorithms. The efficiency of the RANSAC is related to the sample size, the size of data set and the proportion of inliers. The RANSAC is slow when the score of the final of model is low. The non-optimal parameters are often obtained, and the results that are obtained by the objective function may be incorrect when degeneration occurs. According to these limitations, [10] proposed a universal sample consensus (USAC) framework. However, the USAC also adopts minimum set sampling without considering changing the sample size. To improve the data set, [11] proposed the progressive sample consensus (PROSAC). In the PROSAC, a good data set is selected by sorting the initial matching results. Thus, the PROSAC can rapidly obtain the best parameter. Reference [12] proposed the N-Adjacent points sample consensus (NAPSAC) based on the fact that inliers are closer to other inliers than outlies. The performance of the NAPSAC is good when the proportion of inliers is low. However, NAPSAC easily degenerates, and the performance is poor when dealing with data sets whose points are close.

In the sample size, the RANSAC adopt the minimum set sampling method and ignore the fact that the sample size affects the score of the final model [13]. The proportion of inliers is the same as the score of the model in a certain sense. Although the point cloud can be preprocessed to improve the portion of inliers using prior information, such as colour or size, the RANSAC may be very slow when the portion of inliers is small. Therefore, how to obtain a higher score is worth considering. Generally, a higher score represents a better result [14].

In the framework of the sample consensus, the score calculation, which affects the speed of the framework, is an essential part. The score calculation methods, such as straight line, plane, cylinder, and other simple models, are very mature [15]. However, the score calculation of an ellipsoid is not easy. Although RANSAC adopt the distance+normal method to calculate the scores of many models, this method is difficult to apply to the ellipsoid [9], [16]–[18]. The traditional method (distance+normal) that is applied in general quadric surface fitting generally contains two parts: the distance and the normal vector. The candidate is an inlier when satisfying the following two conditions: 1) the minimum distance between the candidate and surface is less than a given value, and 2) the difference between the normal vector of the candidate and the normal vector of the corresponding point on the surface is less than a given value. However, the minimum distance between a candidate point and an ellipsoid is difficult to calculate. In addition, the corresponding points of every candidate point are difficult to confirm. The complicated calculations may cause a slow speed.

The score calculation method is often different when the sample consensus framework applied in different scenarios. When fitting a line and plane, the RANSAC gives a general solution to calculate the score [9]. For example, a candidate point is an inlier of a line when the distance between the candidate and the line is less than the given threshold. Similarly, a candidate is an inlier of a plane when the distance between the candidate and the plane is less than the given threshold. The cylinder is slightly more complicated because of the calculation of the normal vector. The calculation of the distance between the candidate and cylinder is also complicated compared with that of a line and plane. One of the most complicated models is the ellipsoid. At present, very few previous studies address the score calculation of the ellipsoid. The distance+normal method can be applied to the ellipsoid only when there are no demands on the speed of the algorithm.

In this paper, we mainly study the sample consensus framework and the score calculation of the ellipsoid in the context of automatic picking of Lingwu Long Jujubes. The contributions of our work are as follows. First, the similarity between Lingwu Long Jujubes and the ellipsoid model was analysed. On this basis, a solution for Locating jujubes in a natural environment is introduced. After the literature survey [19]–[23], we carefully consider many key details, including multi-targets detection, occlusion problems, and different natural lighting. Besides the position of the jujubes, the size and attitude angle of jujubes in a camera coordinate system for an automatic picking robot is also presented. Second, we propose the variable sample consensus (VARSAC) to accelerate speed of the RANSAC and obtain a best model whose model score is higher than that of RANSAC.

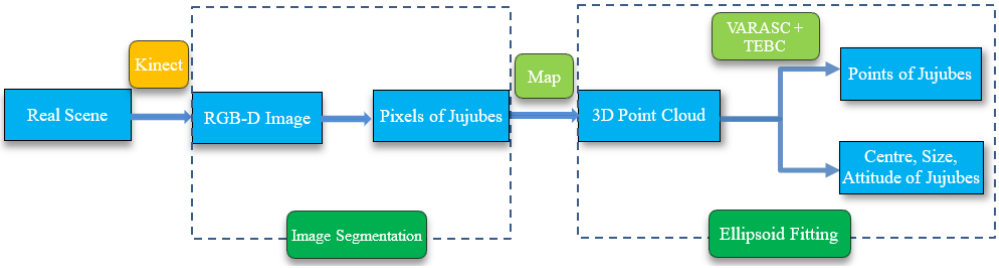

FIGURE 1. Overall process to detect Lingwu Long Jujubes in a real scene.

Meanwhile, the sufficient condition of a higher score is analysed. In VARSAC, we redesign the sample consensus framework and present a new termination criterion of algorithm, which is related with the confidence level. Third, we proposed two-ellipsoid-bounding-counting (TEBC) method, which is simple and visual, to calculate the score of the ellipsoid model. The result shows that the VARSAC+TEBC algorithm has superior performance to the RANSAC+distance+normal algorithm.

II. METHOD

In this section, a method for Locating Lingwu Long Jujubes is proposed. First, the pixels that represent jujubes are extracted using colour threshold. Then, the 3D points are selected using the mapping relation. Finally, the models and the relative parameters are solved by the VARSAC, TEBC and LS methods. The overall process, which includes 2D image segmentation and 3D ellipsoid fitting, is shown in Fig. 1.

A. MODEL HYPOTHESIS OF LINGWU LONG JUJUBES

One of the prior information is that the shape of Lingwu Long Jujubes is similar to an ellipsoid. The prior information needs to be proven before fitting the jujubes. There is a theorem that the shape of an object can be regarded as an ellipsoid when the front projection is an ellipse and bottom is a circle [24]. The theorem will evaluate the similarity between the jujubes and the ellipsoid.

The similarity is calculated using the maximum inscribed ellipse and the maximum inscribed circle. δ describes the level of similarity between the front projection and an ellipse. ρ describes the similarity between the bottom and a circle. Both can be calculated by (1) and (2). δ and ρ are normalized to 0-1. When $\delta \cdot \rho$ is close to 1 or 100%, the ellipsoid model well represents the jujube. S_{fp} and S_{el} respectively denote the areas of the front projection and the ellipse in Fig. 2(a). a_{re} and b_{re} respectively represent the length and width of the rectangle in Fig. 2(b).

$$\delta = \left(1 - \frac{|S_{fp} - S_{el}|}{S_{fp}}\right) \times 100\% \quad (1)$$

$$\rho = \frac{b_{re}}{a_{re}} \times 100\% \quad (2)$$

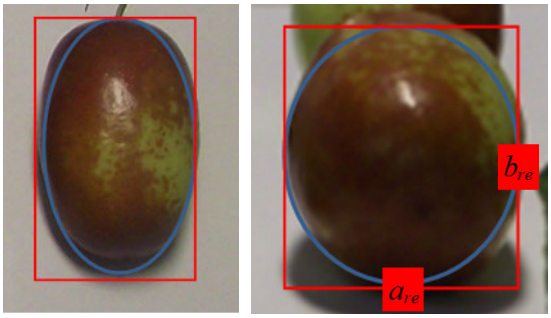

FIGURE 2. Shape of a Lingwu Long Jujubes. (a) image that shows the front projection of a jujube, the minimum bounding rectangle and the maximum inscribed ellipse. (b) image that shows the bottom of the jujube, the minimum bounding rectangle and the maximum inscribed circle.

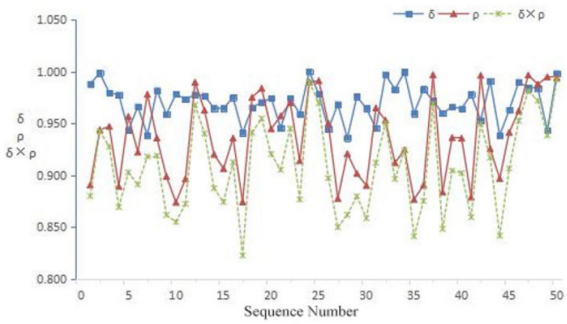

FIGURE 3. The similarity between the Lingwu Long Jujubes and an ellipsoid.

The statistical results of 50 fresh jujubes are shown in Fig. 3. The average values of δ, ρ, and $\delta \cdot \rho$ are 89.2%, 93.7%, and 89.2%, respectively. The results show that the similarity is high and the ellipsoid is a good model for Lingwu Long Jujubes.

As another prior information, the general size of jujubes also is measured. The major axis is 41 mm-46 mm, and minor axis is 20.5 mm-23.2 mm. The size of jujubes will be used to determine whether the model in sample consensus is valid.

B. METHOD OF DETECTING THE JUJUBES

Some background objects may be red-brown or similar to red-brown in different natural lighting. Segmentation using only colour cannot complete our task very well because it

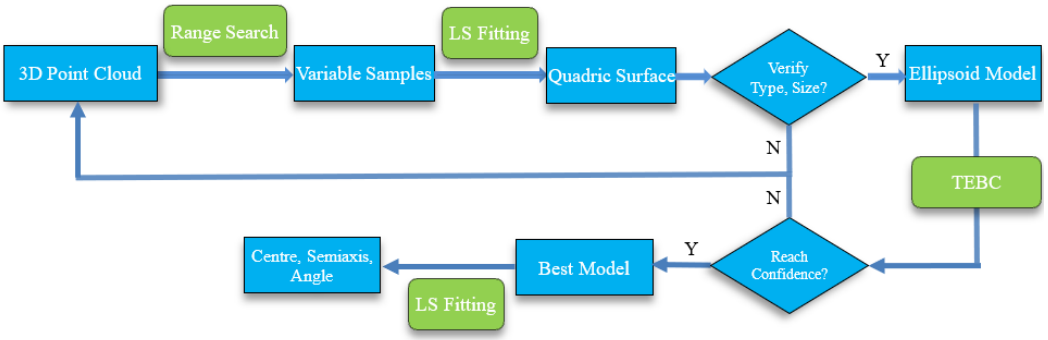

FIGURE 4. Process to detect jujubes in point clouds.

is imprecise. However, image segmentation can preliminarily extract the jujubes from a background. 3D points are produced by depth sensors, and 2D images are produced by CCD cameras. There is a mapping relation between the pixels in a 2D image and 3D points [25]. The pixels, which represent the jujubes, in 2D images can be extracted from the background using the colour threshold. Then, 3D points of the jujubes are selected from point clouds using the mapping relation. In this way, the proportion of inliers will be increased.

There are still some outliers in the image segmentation result. In addition, some points of the jujubes may be not on the fitted ellipsoid because of large errors. Therefore, we considered the LS method and the sample consensus to fit. In the sample consensus framework, the TEBC is used to select inliers and calculate scores. The invariant theory is used to recover the 3D information of jujubes in their natural environment (e.g., centres, sizes and attitude angles of the jujubes).

The multi-targets detection must be considered because there may be multiple jujubes in the same field of view. The time that is spent detecting the first jujube may be very large because the proportion of the points, which belong to the first jujube, is low. In the sampling stage, two points from different jujubes could be sampled at the same time. Thus, the amount of invalid iterative process increases. The pristine RANSAC framework cannot handle multi-targets detection. Considering these factors, we mainly use the range search method of the K-D tree to solve the multi-targets detection problem. An appropriate radius r_s of the range search is set to avoid sampling points of different jujubes.

The process of 3D point cloud is shown in Fig. 4. The specific steps of our solution to detect multiple jujubes are as follows.

Step 1. image segmentation using colour.

The original point cloud is obtained by a depth sensor. Then, the colour threshold in the HSV space is calculated, and the points of the jujubes are preliminarily selected by the mapping relation.

Step 2. Range search with K-D tree.

Create a K-D tree for the point cloud and randomly sample one point as the initial seed point. The other nearest points to the seed point are obtained by K-D tree whose search range is set according to the general size of the jujubes.

Step 3. Fit the model using the LS, determine the model type and verify the model by comparing the general size of the jujubes and the semiaxis of the model.

We can obtain the quadric surface equation using the LS method and determine whether the model is an ellipsoid using the invariants of the quadric surface. If this model is the ellipsoid, the semiaxis of the ellipsoid will be calculated to verify again. Then, the best model is determined according to their scores, and we proceed to *Step 4*. Otherwise, we proceed to *Step 2*.

Step 4. Extract the best ellipsoid.

The inliers in the best ellipsoid are extracted from the point cloud. The coefficients of the general equation are known from the returned model. The sizes, centres and attitude angles of the jujubes are calculated using the coefficients of the general equation.

Step 5. Detect other jujubes in the remaining points.

The remaining points will be subjected to *Step 2* until the number of the remaining points is less than the number of inliers in the last best ellipsoid.

The principle of the ellipsoid fitting is as follows. The quadric equation is defined as (3).

$$0 = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + a_{14}x + a_{24}y + a_{34}z + a_{44} + a_{12}xy + a_{13}xz + a_{23}yz \quad (3)$$

The quadratic coefficient a_{11} must not be 0 for an ellipsoid. Therefore, (3) can be reformed as (4) and $a_{11} = 1$.

$$0 = x^2 + a_{22}y^2 + a_{33}z^2 + a_{14}x + a_{24}y + a_{34}z + a_{44} + a_{12}xy + a_{13}xz + a_{23}yz \quad (4)$$

There are 9 parameters to be solved in (4). Thus, at least 9 points are sampled for sample consensus framework and the error e_i of the i-th sample point (x_i, y_i, z_i) is defined in (5).

$$e_i = x_i^2 + a_{22}y_i^2 + a_{33}z_i^2 + a_{14}x_i + a_{24}y_i + a_{34}z_i + a_{44} + a_{12}x_iy_i + a_{13}x_iz_i + a_{23}y_iz_i \quad (5)$$

The objective function of the optimizing search is $\sum e_i^2$. The process can be described as the solution of the linear equation $Ax = v$, where A, v is the known matrix that can be obtained by the partial derivative. x denotes the unknown parameters of quadric surface equation and $x = [a_{22}, a_{33}, a_{14}, a_{24}, a_{34}, a_{44}, a_{12}, a_{13}, a_{23}]^T$.

The invariants I_1, I_2, I_3, and I_4 of the quadric surface can be calculated by (6), (7), (8), and (9), respectively. And then, the type of quadric surface is determined by the invariants. A sufficient condition that a quadric surface is an ellipsoid is that I_1, I_2, and I_3 are greater than 0 [24].

$$I_1 = a_{11} + a_{22} + a_{33} \tag{6}$$

$$I_2 = \begin{vmatrix} a_{11} & a_{12}/2 \\ a_{12}/2 & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{13}/2 \\ a_{13}/2 & a_{33} \end{vmatrix} + \begin{vmatrix} a_{22} & a_{23}/2 \\ a_{23}/2 & a_{33} \end{vmatrix} \tag{7}$$

$$I_3 = \begin{vmatrix} a_{11} & a_{12}/2 & a_{13}/2 \\ a_{12}/2 & a_{22} & a_{23}/2 \\ a_{13}/2 & a_{23}/2 & a_{33} \end{vmatrix} \tag{8}$$

$$I_4 = \begin{vmatrix} a_{11} & a_{12}/2 & a_{13}/2 & a_{14}/2 \\ a_{12}/2 & a_{22} & a_{23}/2 & a_{24}/2 \\ a_{13}/2 & a_{23}/2 & a_{33} & a_{34}/2 \\ a_{14}/2 & a_{24}/2 & a_{34}/2 & a_{44} \end{vmatrix} \tag{9}$$

1) THE POSITION OF THE ELLIPSOID

According to spatial geometry theorem, the centre coordinate of a quadric surface is the solution of the linear equations in (10) [24].

$$\begin{bmatrix} a_{11} & a_{12}/2 & a_{13}/2 \\ a_{12}/2 & a_{22} & a_{23}/2 \\ a_{13}/2 & a_{23}/2 & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -a_{14}/2 \\ -a_{24}/2 \\ -a_{34}/2 \end{bmatrix} \tag{10}$$

2) THE ATTITUDE ANGEL OF THE ELLIPSOID

$A_{3\times3}$ is the coefficient matrix on the left side of (10), and its characteristic roots are the solution of (11) [24].

$$-\lambda^3 + I_1 \cdot \lambda^2 - I_2\lambda + I_3 = 0 \tag{11}$$

There must be three characteristic roots $\lambda_1, \lambda_2, \lambda_3$ and three main directions v_1, v_2, v_3. The main directions respectively represent the direction vectors of the semiaxis of an ellipsoid. Then, the rotation angles θ_x, θ_y, θ_z of an ellipsoid can be calculated using the rotation matrix consisting of v_1, v_2, v_3.

3) THE SIZE OF THE ELLIPSOID

The standard form of an ellipsoid is shown in (12). Thus, the semiaxis a, b, and c can be calculated by (13) [24], [26].

$$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + I_4/I_3 = 0 \tag{12}$$

$$\begin{cases} a^2 = -\lambda_1^{-1} \cdot I_4/I_3 \\ b^2 = -\lambda_2^{-1} \cdot I_4/I_3 \\ c^2 = -\lambda_3^{-1} \cdot I_4/I_3 \end{cases} \tag{13}$$

C. SAMPLE CONSENSUS – VARIABLE SAMPLE SIZE

In this section, the reason why the variable sample size method is better than the minimum set sampling method will be discussed. Then, the concrete steps of the VARSAC will be given.

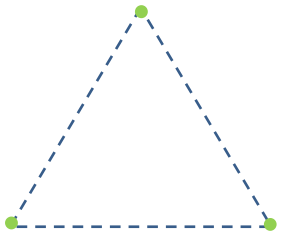

FIGURE 5. The fitting of three points using the different sample size.

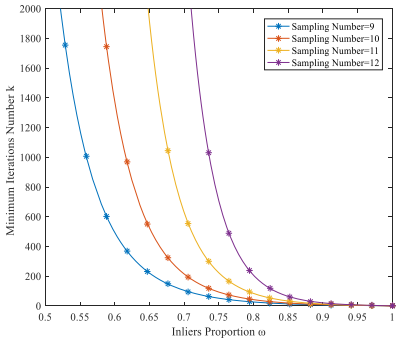

FIGURE 6. The number of iterations changes with the proportion of inliers and the sample size.

The iterations k of the sample consensus can be calculated by (14) [9].

$$1 - p = (1 - w^n)^k \tag{14}$$

where p denotes the confidence level. n is the sample size. w is the proportion of inliers, and it changes with n. (14) can be reformed as (15).

$$k = log\,(1 - p)\,/log\left(1 - w^n\right) \tag{15}$$

p is constant and needs to be set in advance. $\partial k/\partial w < 0$ when $0 < w < 1$. $\partial k/\partial n > 0$ when $0 < n$. We desire that k be small to get a faster algorithm. Therefore, we hope that n is small, and w is large. Unfortunately, w changes with n, and this change is unpredictable. For example, as shown in Fig. 5, the data set is the three vertexes of the regular triangle, and we want to find a best straight line using the sample consensus for the three points. We may set a distance threshold so that the inliers are the three vertexes ($w = 1$) when $n = 3$. We also may set a distance threshold so that the inliers are two of the vertexes ($w = 2/3$) when $n = 2$. According to (15), k will be 7 when $n = 2$, $p = 0.98$ and k will be 1 when $n = 3$, $p = 0.98$. Thus, the fitting of $n = 3$ is faster than the fitting of $n = 2$. This example proves that w is different when the sample size n is changed. The value of n affects w and then affects k. Thus, we want to find a higher score by including some unequal sample sizes to accelerate the algorithm.

For the ellipsoid fitting, Fig. 6 shows the functional relationship between k and w when the values of n are different. k decreases as n decreases when w is constant. k decreases

TABLE 1. VARSAC for the fitting of the ellipsoid.

Algorithm 1: Variable Sample Consensus

Input: n_m (m sample sizes), p , k_{max}

Output: M_{best} , k

i=0 , j=0 , s_{max}=0

while $i < k$ and $i < k_{max}$ **do**

  while $j < m$ **do**

    1. Model Estimate

    i++, j++

    Randomly sample subset of n_j points

    Estimate model parameters M_i

    2. Verification

    Verify M_i with size and invariants

    Calculate the score s_i of M_i by TEBC

    if $s_i > s_{max}$ **then**

      M_{best}=M_i , s_{max}=s_i

      Recompute k from (17) using $w_j = s_{max}$

    else

      abandon M_i

    end if

  end while

  j=0

end while

optimization M_{best} with the inner point

calculate the size、 angle and centre of M_{best}

M_{best} denotes the best model, and M_i denotes the model of the i -th iteration. s_{max} denotes the score of the best model. k_{max} denotes the upper limit on the number of iterations.

when w increases when n is constant. k may increase when we abandon minimum set sampling. At this moment, there is an inevitable risk. The time costs of the algorithm will increase compared with those of minimum set sampling if the maximum of w cannot be found as soon as possible or it is not large enough. The different sample size result in the different score. Therefore, we try to find the maximum of w by including some unequal sample sizes to accelerate the algorithm.

Based on the above discussion, we proposed a novel sample consensus framework (see Table 1). The VARSAC traverses n to reduce the time costs with a high probability. The condition that the VARSAC is faster than the RANSAC will be discussed as follows.

$$1 - p = \left(\prod_{j=1}^{m} \left(1 - w_j^{n_j}\right)\right)^{k/m} \tag{16}$$

(16) can be derived from (15). The minimum set sampling is adopted in the RANSAC. Thus, the VARSAC is the RANSAC when $m = 1$. And then, (16) can be reformed as (17). The actual number of iterations k of VARSAC can be calculated by (17).

$$k = m \times log\,(1-p)/log\left(\prod_{j=1}^{m} 1 - w_j^{n_j}\right) \tag{17}$$

The VARSAC has fewer iterations k than the minimum set sampling that is adopted in the RANSAC as long as $\prod_{j=1}^{m}\left(1 - w_j^{n_j}\right) < 1 - w_1^{n_1}$.

$$k = log\,(1-p)/\left(\sum_{j=1}^{m} log\left(1 - w_j^{n_j}\right)/m\right) \tag{18}$$

(17) can be reformed as (18). Both of these values have the same numerator, and k is determined by $\sum_{j=1}^{m} log\left(1 - w_j^{n_j}\right)/m$. Formally, $\sum_{j=1}^{m} log\left(1 - w_j^{n_j}\right)/m$ for the VARSAC looks like the average of a group of numbers. $\sum_{j=1}^{m} log\left(1 - w_j^{n_j}\right)/m$ where $m = 1$ for the RANSAC looks like the first one of this group of numbers. Then, whether the VARSAC is faster than the RANSAC is determined by the features of this group of numbers. $log\left(1 - w_j^{n_j}\right) < 0, j = 1, 2, \ldots, m.$ is noticed and (19) is true statement. Then, the sufficient condition for the VARSAC to be faster than the RANSAC is (20).

$$\begin{aligned} &-\sum_{j=1}^{m} log\left(1 - w_j^{n_j}\right)/m \\ &> -log\left(1 - max\left\{w_j^{n_j}, j = 1, 2, \ldots, m.\right\}\right)/m \qquad (19) \\ &\quad -log\left(1 - max\left\{w_j^{n_j}, j = 1, 2, \ldots, m.\right\}\right)/m \\ &> -log\left(1 - w_1^{n_1}\right) \qquad (20) \end{aligned}$$

A new function is defined in (21) to compare the two values in (20). Then, (22) is the total differential equation of (21).

$$z = -log\,(1-t)/q \tag{21}$$

$$\begin{aligned} \Delta z \approx\ & -1/\left(q \times ln\,(10) \times (t-1)\right) \times \Delta t \\ & + log\,(1\text{ -}t)/\left(q^2 \times ln\,(10)\right) \times \Delta q \end{aligned} \tag{22}$$

where $\Delta z = z\,(t_2, q_2) - z\,(t_1, q_1)$ when $q_1 = 1$, $t_1 = w_1^{n_1}$, $q_2 = m$, and $t_2 = max\left\{w_j^{n_j}, j = 1, 2, \ldots, m.\right\}$. $\Delta z > 0$ shows that the VARSAC is faster than the RANSAC.

According to (21) and (22), we find that one of the sufficient conditions of $\Delta z > 0$ for our ellipsoid fitting is that $n_m = \{9, 10, 11, 12\}$ and $w_1 < max\left\{w_j, j = 1, 2, \ldots, 4.\right\} - 0.2$.

D. TWO-ELLIPSOID-BOUNDING COUNTING

Currently, there are few score calculation methods for the ellipsoid model. If we apply distance+noraml method to an ellipsoid, we need to calculate the distance between each candidate and the ellipsoid. The normal vectors of each candidate and the corresponding point in the ellipsoid also need to be calculated. This method is complex and time-consuming when applied to ellipsoids. In this section, we propose the TEBC to calculate the score of the ellipsoid model.

First, two auxiliary ellipsoids, which include a large ellipsoid and a small ellipsoid, are produced using the method of scaling the semiaxis of the fitted ellipsoid in each iteration.

Additionally, the auxiliary ellipsoids have the same position and attitude with the fitted ellipsoid. The steps are as follows.

1) OBTAIN THE GENERAL EQUATIONS OF the AUXILIARY ELLIPSOIDS USING THE FITTED ELLIPSOID

The standard equation of the fitted ellipsoid is (23).

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1 \tag{23}$$

The standard equations of two auxiliary ellipsoids corresponding to the fitted ellipsoid are (24).

$$\frac{x^2}{(sa)^2}+\frac{y^2}{(sb)^2}+\frac{z^2}{(sc)^2}=1 \tag{24}$$

where s is the scaling factor. $s = 1$ denotes the standard equation of the fitted ellipsoid. $s > 1$ denotes the standard equation of the large ellipsoid. $0 < s < 1$ denotes the standard equation of the small ellipsoid.

The general equation of the fitted ellipsoid is known. The general equations of the auxiliary ellipsoids will be obtained by the general equation of the fitted ellipsoid. The relationship of the general equation between the auxiliary ellipsoids and the fitted ellipsoid is discussed as follows.

First, we assumed that the process of the coordinate transformation from the standard equation to the general equation is the same for the three ellipsoids (the fitted ellipsoid and two auxiliary ellipsoids). Then, the transformation is described by the rotation matrix R and translation matrix T. The results are different if the orders of the transformation are different. Generally, rotation and then the translation is adopted. The relationship of the coordinate (x', y', z') after the transformation and the coordinate (x, y, z) before the transformation is shown in (25).

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + T \tag{25}$$

Then, we substitute (25) into (24) to obtain the general equations of the three ellipsoids, as shown in (26).

$$f\left(x', y', z'\right) = s^2 \tag{26}$$

Finally, we find an important detail based on (26). The corresponding coefficients of the three ellipsoids are equal except for the constant a_{44}. In other words, the general equations of the auxiliary ellipsoids can be obtained by only changing the constant a_{44} of the fitted ellipsoid. In addition, we also find following inferences according to the above discussion. i). The invariants I_1, I_2, and I_3 of the three ellipsoids are equal but I_4 is not equal. ii). The characteristic equations and the characteristic roots λ_1, λ_2, and λ_3 of the three ellipsoids are the same because the characteristic equations are only related to I_1, I_2, and I_3.iii). According to (13), the relationship between I_4'' and I_4' is shown in (27). The invariant of the auxiliary ellipsoids can be calculated by (27).

$$I_4'' = s^2 I_4' \tag{27}$$

where I_4'' denotes I_4 of the auxiliary ellipsoid, and I_4' denotes I_4 of the fitted ellipsoid. The constant a_{44} of the two auxiliary ellipsoid can be calculated by I_4'', and then the general equations of the auxiliary ellipsoids are obtained.

2) CALCULATE THE NUMBER OF POINTS BETWEEN THE LARGE ELLIPSOID AND THE SMALL ELLIPSOID

An ellipsoid is a closed surface, and we can divide the points into three kinds based on their spatial location: points on the ellipsoid, inside points and outside points. $F(x, y, z) = f(x, y, z) - 1$ is defined as the general equation of an ellipsoid. P is assumed to be a candidate. The position relationship between the candidate and the ellipsoid is determined by the relationship between $F\left(x_p, y_p, z_p\right)$ and 0.

i). P is on the ellipsoid. Thus, $f\left(x_p, y_p, z_p\right) = 1$ and (28) holds.

$$F\left(x_p, y_p, z_p\right) = 0 \tag{28}$$

ii). P is inside the ellipsoid. P must be on the certain small ellipsoid. Thus, $f\left(x_p, y_p, z_p\right) = s^2$, where $0 < s < 1$. Then, $f\left(x_p, y_p, z_p\right) < 1$, and (29) holds.

$$F\left(x_p, y_p, z_p\right) < 0 \tag{29}$$

iii). P is outside the ellipsoid. P must be on the certain large ellipsoid. Thus, $f\left(x_p, y_p, z_p\right) = s^2$, where $s > 1$. Then, $f\left(x_p, y_p, z_p\right) > 1$, and (30) holds.

$$F\left(x_p, y_p, z_p\right) > 0 \tag{30}$$

In TEBC, a candidate is regarded as an inlier of the fitted ellipsoid when the candidate is inside the large ellipsoid and outside the small ellipsoid. According to (28), (29), and (30), we can easily find the position relationship between a candidate and the auxiliary ellipsoids, and then determine whether a candidate is an inlier or not.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Many experiments with simulation data and real dataset have been conducted to validate the proposed method. The simulation data are mainly used to show the process of VARSAC+TEBC and the error of every point. The real dataset are mainly used to test the accuracy and speed in the natural environment. The method of direct fitting is also included in the experiments for comparison. The sample size of the VARSAC is set as 9, 10, 11, and 12. The sample size of the RANSAC is set as 9 because the RANSAC adopts the minimum set sampling method. The distance+normal is adopted by the RANSAC while the TEBC is adopted by the VARSAC.

The different methods are evaluated by the following criteria. The model score is in the form of the proportion, and the algorithm is fast when the score is high. RMS-s, RMS-c and RMS-a respectively denote the accuracies of size, centre and attitude angle.

i). Model score: $\frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{total_i} \times 100\%$.

ii). RMS-s (Root mean squared error of the size):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{3}\sum_{\ell_i\in a_i,b_i,c_i}\left|\ell_i^*-\ell_i\right|\right)^2}.$$

iii). RMS-c (Root mean squared error of the centre):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{3}\sum_{\ell_i\in c_{xi},c_{yi},c_{zi}}\left|\ell_i^*-\ell_i\right|\right)^2}.$$

iv). RMS-a (Root mean squared error of the attitude angle):

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{3}\sum_{\ell_i\in \theta_{xi},\theta_{yi},\theta_{zi}}\left|\ell_i^*-\ell_i\right|\right)^2}.$$

where the index i denotes the i-th fitted ellipsoid. n denotes the number of fitted ellipsoids in the data set. m_i denotes the score of the best ellipsoid. $total_i$ denotes the total number of points before fitting. ℓ_i^* and ℓ_i denote the ground-truth value and the fitted value, respectively.

Some experimental parameters need to be set in advance. The confidence level p is usually set to 0.95-0.99, and 0.95 is adopted in our experiment to achieve faster fitting. The scaling factor s is the parameter of the TEBC that includes two values corresponding to two auxiliary ellipsoids. For convenience, we just need to set a variable portion d $(1 \pm d)$ for scaling factor s, such as 1 ± 0.1, 1 ± 0.2 and so on. The variable portion d is value between 0 and 1. More points are regarded as inliers when d is larger. The sampling radius r_s is set to avoid sampling two points from the different ellipsoids. According to the size of jujubes, the minor semiaxis and major semiaxis are set to verify whether the fitted ellipsoid conforms to the prior knowledge or not. The upper limit k_{max} on the number of iterations is set in advance. It will work to avoid waste of time if the current best model not reach the confidence level and too much time is consumed.

Throughout the experiment, the score calculation method of the VARSAC is the TEBC and that of the RANSAC is the distance+normal.

A. EXPERIMENTS ON SIMULATION DATA.

The first experiment was carried out on the simulation data set. The parameter initialization for the VARSAC is shown in Table 2. In the simulation data, there is an ellipsoid whose equation is known and some random noise, as shown in Fig. 7(a). The gaussian distribution errors are added in the points on the ellipsoid. The portion of random noise is approximately 30%, and some noise may become inliers because it may be on the ellipsoid. The simulation data can accurately analyse the error of every point and the errors of the fitted parameters because there are not measuring the errors of the real data.

Fig. 7 shows the result of the simulation data in different stages. Fig. 7(b) shows two auxiliary ellipsoids of the TEBC and the process of removing most of the random noise. The point cloud in the best model is shown in Fig. 7(c) and the final fitted ellipsoid that is drawn by the fitted equation is shown in Fig. 7(d).

TABLE 2. Parameter initialization for the VARSAC.

Symbol	Value
p - confidence level	0.95
s -scaling factor	1 ± 0.1
r_s -sampling radius	1
min radius	0.5
max radius	4
k_{max}	3000

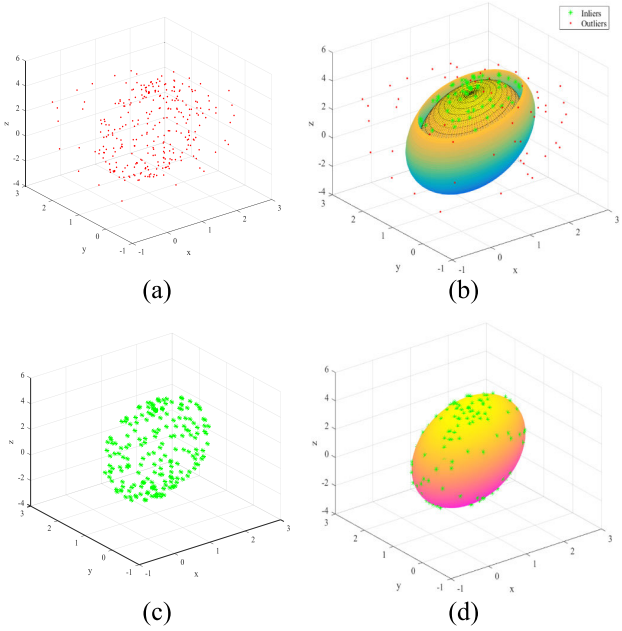

FIGURE 7. The experimental data and the results. (a) Image that shows the simulation data including an ellipsoid and random noise. (b) Image that shows the score calculation by the TEBC. (c) Image that shows the inliers that were obtained by the VARSAC. (d) Image that shows the final result.

Table 3 shows the fitted parameters including the sizes, centres and attitude angles. The attitude angles θ_x, θ_y, and θ_z represent the rotation matrix in the transformation. The relative errors of attitude angles are not shown because the true value may be 0, and 0 cannot be the denominator. The relative error of the fitted parameters is small and the maximum is not more than 4.55%.

We test different methods on the simulation data. The evaluation is shown in Table 4. The time consumption of the score calculation is stable. The average time consumption of the TEBC method is 2 ms while that of the distance+normal method is 5 ms. The time consumption of an iteration is 0 ms if the samples in this iteration not support an ellipsoid.

We also analyse the error of every inlier and the error of the i-th inlier (x_i, y_i, z_i) is defined as $F(x_i, y_i, z_i)$. The Fig. 8 contrasts the performance of the RANSAC and VARSAC using the known equation.

In Figs. 8(a) and 8(b), the error distributions of the different methods are different, and the error distribution of the VARSAC is better than that of the RANSAC. The error

TABLE 3. The relative parameters of the fitted ellipsoid.

Parameter	a	b	c	c_x	c_y	c_z	θ_x /rad	θ_y /rad	θ_z /rad
Ground truth	1	2	3	1	1	1	0	0	1.05
Fitted value	1.0098	2.0354	3.0462	0.9647	0.9740	0.9545	0.11	-0.10	1.13
Relative error	0.98%	1.77%	1.54%	3.53%	2.60%	4.55%	---	---	---

'---' represent incalculable criteria.

TABLE 4. Evaluation of different methods using the simulation data.

Method	Score	RMS-c	RMS-s	RMS-a/rad	Iteration	Time1/ms	Time2/ms
Direct fitting [6]	**100%**	0.54	0.61	0.87	**1**	**13**	**0**
RANSAC [9]	61.2%	0.07	0.08	0.43	282	2002	5
Ours	78.3%	**0.02**	**0.02**	**0.11**	133	505	2

Time1 represent the time consumption of the whole process. Time2 represent the average time consumption of the score calculation.

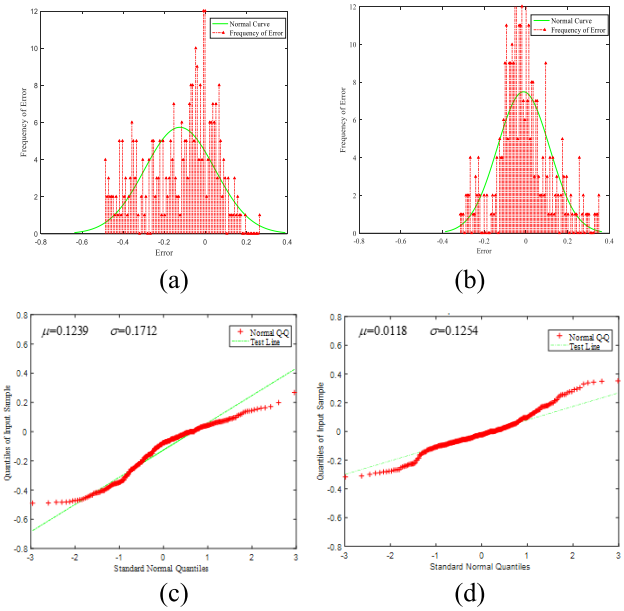

FIGURE 8. The error analysis and comparison. (a) image that shows inlier errors of the RANSAC. (b) image that shows the inlier errors of the VARSAC. (c) image that presents the Q-Q Plot of the RANSAC. (d) image that presents the Q-Q Plot of the VARSAC.

distribution is approximate to the normal distribution. Therefore, the inlier error is analysed by the Q-Q Plots that are shown in Figs. 8(c) and 8(d). The mean is 0.1239, and the standard deviation is 0.1712 for the RANSAC. The mean is 0.0118, and the standard deviation is 0.1254 for the VARSAC. Compared with the RANSAC, the VARSAC improves the error of every inliers.

B. EXPERIMENTS IN NATURAL ENVIRONMENT

The second experiment mainly includes the following parts. First, the jujubes are extracted from the background using the colour threshold in the 2D image. Second, the jujube, which is similar to ellipsoid in 3D points, is detected using the sample consensus.

The Kinect v2 can obtain more points than Kinect v1 in bright light and is more accurate than the Kinect v1 by approximately 20 pp at a close range (less than 4 m). Therefore, we collected the data using the Kinect v2 in Lingwu county, Ningxia Province. The specific parameters are as follows: image size of 512×424, depth sensor accuracy of 380μm at close distance, and shooting distance of 0.5 m-4 m [27].

1) IMAGE SEGMENTATION

The experiment to determine the threshold is performed using the 2D images that were taken by the colour camera of Kinect v2. The illumination in different regions changes significantly with time, especially in this agriculture environment [28], [29]. The images being in the HSV colour space make it easier to distinguish jujubes from background. H, which represents the hue of an object, is not sensitive to the illumination [30], [31]. Therefore, H is used to conduct image segmentation. According to the distribution statistics of colour, the threshold of H is set to 0-36 or 252-360 in the experiment.

Some pixels of the jujube border may be removed in image segmentation. Thus, we use mean filter to prevent from losing the edge pixels of the jujube in image segmentation. For further reduce the outliers in Fig. 9(c), we set the region of interest (ROI) based on the area of the connected region. The connected region whose area is small will be removed, as shown in Fig. 9(d). The final result contains few pixels that represent the background. The proportion of inliers after image segmentation is favourable for the sample consensus.

2) POINT CLOUD SEGMENTATION

After 2D image segmentation, the 3D points that were collected by the depth sensor are selected according to the mapping relation between the pixels in a 2D image and 3D points. As shown in Fig. 10(a), the original point cloud is obtained.

The parameter initialization of the algorithm is shown in Table 5. According to the general sizes of Lingwu Long

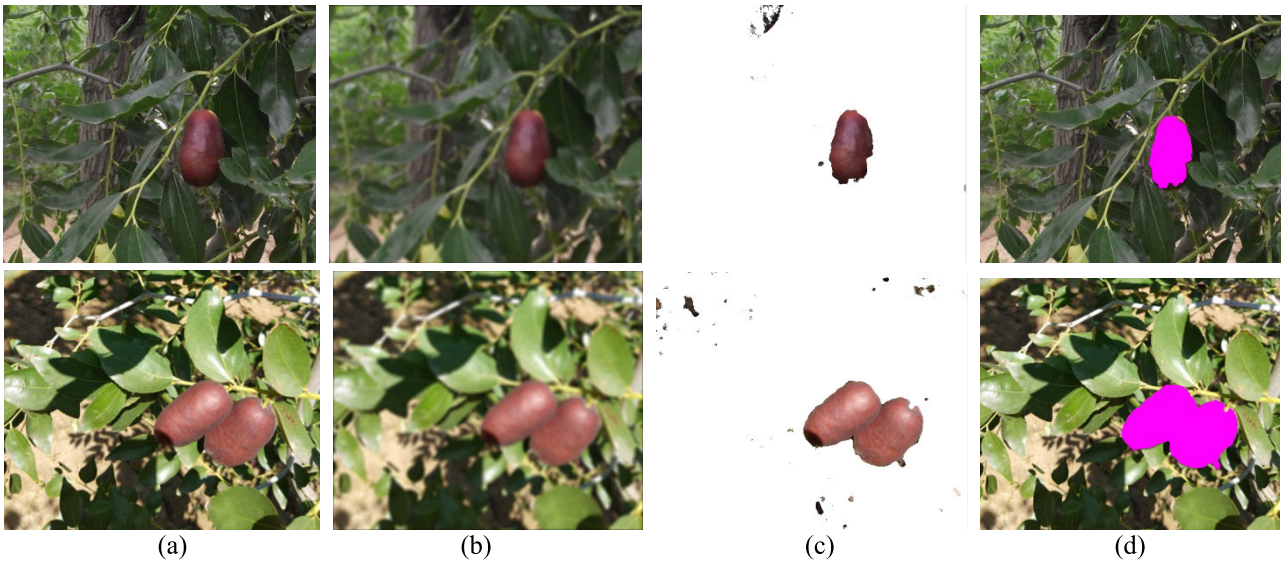

FIGURE 9. Image segmentation in the HSV space. (a) RGB images. (b) images after mean filtering. (c) images after image segmentation. (d) sets ROI and removes the connected region whose area is small.

TABLE 5. Parameters initialization for VARSAC.

Symbol	Value
p - confidence level	0.95
s -scaling factor	1 ± 0.1
r_s -sampling radius	30 mm
minor semiaxis	10 mm
major semiaxis	23 mm
k_{max}	3000

Jujubes, the semiaxis of the ellipsoid to be detected is limited to 10 mm-23 mm. The sample range of the K-D tree is set as 30 mm according to space between two jujubes. The algorithm will terminate when the number of iterations reaches the upper limit k_{max}.

In the experiment of the real dataset, we also apply the multi-targets detection strategy to the RANSAC for the contrast experiment. In Figs. 10(b) and 10(c), two jujubes, which distinguished by different colours, are successfully detected. The number of the points represents the score. From Fig. 10, we can find that the VARSAC get a higher score than the RANSAC.

All of the single jujubes are selected to test the performances of the different method in the real data set and we took the averages in the evaluation.

In Table 6, the VARASAC has a higher score of approximately 77.4% compared to the 60.3% of the RANSAC. The method is faster when its score is higher. The total time consumption of the direct fitting is the shortest because this method does not need to calculate the score of the model, and it only needs one iteration. According to total time consumption, our method takes approximately 588 ms while the RANSAC takes approximately 2310 ms. The speed of the

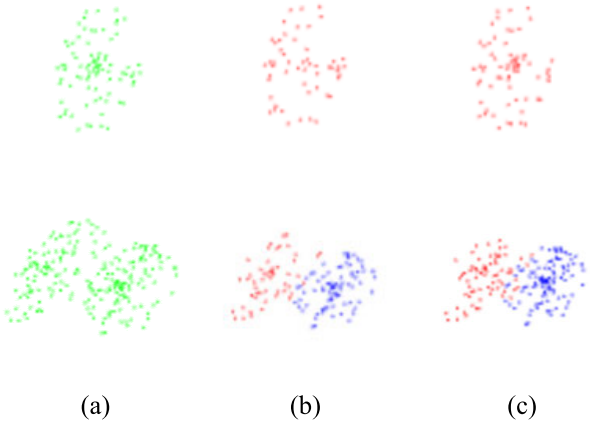

FIGURE 10. Point cloud segmentation by the RANSAC and VARSAC. (a) original point clouds from image segmentation. (b) segmentation by the RANSAC. (c) segmentation by the VARSAC.

VARSAC+TEBC is approximately 4 times faster than that of the RANSAC.

The VARSAC try to find the highest score in the different sample sizes that includes the minimum set sampling. Meanwhile, the RANSAC only try to find the highest score in the minimum sample set. Because the randomness of sampling, the VARSAC has the risk of failure to obtain a higher score than the RANSAC. So, we find that the VARSAC may be slower than the RANSAC when the score of the final model is not large enough. The risk of failure is low in the whole experiment because the inliers are more likely to be sampled when their corresponding score is higher.

In the score calculation, the average time consumption per valid iteration (Time2) is used to evaluate the score calculation speed. TEBC is 2.5 times as fast as the distance+normal

TABLE 6. Evaluation of different methods.

Method	Score	RMS-*c*/mm	RMS-*s*/mm	RMS-*a*/rad	Iteration	Time1/ms	Time2/ms
Direct fitting [6]	**100%**	4.34	7.13	0.53	**1**	**11**	**0**
RANSAC [9]	60.3%	2.09	2.96	0.50	303	2310	5
Ours	77.4%	**1.80**	**1.21**	**0.46**	147	588	2

Time1 represent the time consumption of the whole process. Time2 represent the average time consumption of the score calculation.

method and both the time consumption are relative stable in the real dataset. This simple calculation contributes significantly to decreasing the total time consumption of the VARSAC.

Just like the distance threshold of the distance+normal method, the accuracy of the VARSAC is controlled by the scaling factor of the TEBC. Table 6 also shows the RMSEs of the different methods with respect to position, size, and attitude angle. With respect to accuracy, our method is better than the RANSAC and Direct fitting. Unfortunately, the RMS-*a* is approximate value because the manual measurement of the attitude angle is difficult in the real world. The accuracy of Direct fitting is the lowest, probably because its results contain more points with large errors. For our task of picking Lingwu Long Jujubes, the accuracies of 1.80 mm in RMS-*p*, 1.21 mm in RMS-*s* and 0.46 rad in RMS-*a* are satisfactory for automatic picking.

IV. CONCLUSION

Locating jujubes in a natural environment is a challenging task. In this paper, the fitting of the ellipsoid is used to locate jujubes. In the sample consensus, the score of different sample size is different. Based on this fact, the VARSAC is introduced to obtain a higher score than the RANSAC. On the one hand, inliers are more likely to be sampled when the score is high. On the other hand, the number of iterations can be decreased. To solve the ellipsoid score calculation problem, the TEBC is proposed. The experimental results show that our method is more accurate and faster than the RANSAC+distance+normal. Although the advantages of the VARSAC are theoretically and experimentally proven, the VARSAC has the risk of failure to obtain a higher score than the RANSAC because of the randomness of sampling. However, this is just a rare phenomenon in the experiment because the inliers are more likely to be sampled when their corresponding score is higher.

In the score calculation, we have found a better expansion of the TEBC (not only for the ellipsoid). In the future, we will redesign and expand the bounding count method to other complex quadric surfaces. We will also apply our method to other useful applications, such as 3D SLAM, image matching or surface fitting.

REFERENCES

[1] J. P. Wachs, H. I. Stern, T. Burks, and V. Alchanatis, "Low and high-level visual feature-based apple detection from multi-modal images," *Precis. Agricult.*, vol. 11, no. 6, pp. 717–735, Dec. 2010.
[2] W. Jia, D. Zhao, X. Liu, S. Tang, C. Ruan, and W. Ji, "Apple recognition based on *K*-means and GA-RBF-LMS neural network applicated in harvesting robot," *Trans. Chin Soc. Agric. Eng.*, vol. 31, no. 18, pp. 175–183, Sep. 2015.
[3] R. K. Megalingam, G. V. Vivek, S. Bandyopadhyay, and M. J. Rahi, "Robotic arm design, development and control for agriculture applications," in *Proc. 4th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Coimbatore, India, Jan. 2017, pp. 1–7.
[4] O. R. Chum and J. Matas, "Optimal randomized RANSAC," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 8, pp. 1472–1482, Aug. 2008.
[5] M. Kesäniemi and K. Virtanen, "Direct least square fitting of hyperellipsoids," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 63–76, Jan. 2018.
[6] X. Ying, L. Yang, J. Kong, Y. Hou, S. Guan, and H. Zha, "Direct least square fitting of ellipsoids," in *Proc. 21st Int. Conf. Pattern Recognit.*, Tsukuba, Japan, Nov. 2012, pp. 3228–3231.
[7] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, May 1999.
[8] Q. Li and J. G. Griffiths, "Least squares ellipsoid specific fitting," in *Proc. Geom. Modeling Process.*, Beijing, China, Apr. 2004, pp. 335–340.
[9] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[10] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm, "USAC: A universal framework for random sample consensus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 2022–2038, Aug. 2013.
[11] O. Chum and J. Matas, "Matching with PROSAC—Progressive sample consensus," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 220–226.
[12] D. R. Myatt, P. H. S. Torr, S. J. Nasuto, J. M. Bishop, and R. Craddock, "NAPSAC: High noise, high dimensional robust estimation," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 458–467.
[13] R. Raguram, J. Frahm, and M. Pollefeys, "Exploiting uncertainty in random sample consensus," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep./Oct. 2009, pp. 2074–2081.
[14] R. Raguram, J.-M. Frahm, and M. Pollefeys, "A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 500–513.
[15] T. M. Awwad, Q. Zhu, Z. Du, and Y. Zhang, "An improved segmentation approach for planar surfaces from unstructured 3D point clouds," *Photogramm. Rec.*, vol. 25, no. 129, pp. 5–23, Mar. 2010.
[16] G. Taubin, "Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 11, pp. 1115–1138, Nov. 1991.
[17] J. Yu, H. Zheng, S. R. Kulkarni, and H. V. Poor, "Outlier elimination for robust ellipse and ellipsoid fitting," in *Proc. 3rd IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process. (CAMSAP)*, Aruba, The Netherlands, Dec. 2009, pp. 33–36.
[18] F. Mai, Y. S. Hung, H. Zhong, and W. F. Sze, "A hierarchical approach for fast and robust ellipse extraction," *Pattern Recognit.*, vol. 41, no. 8, pp. 2512–2524, Aug. 2008.
[19] A. Mian, "3D model-based object recognition and segmentation in cluttered scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1584–1601, Oct. 2006.
[20] C. Jia, F. Shi, Y. Zhao, Z. Wang, M. Zhao, and S. Chen, "Identification of pedestrians from confused planar objects using light field imaging," *IEEE Access*, vol. 6, pp. 39375–39384, 2018.
[21] Point Cloud Library. (2017). *Point Cloud Library: The Standalone, Large Scale, Open Project for 2D/3D Image and Point Cloud Processing*. [Online]. Available: http://www.pointclouds.org/

[22] J. Lu and X. Hu, "Detecting green citrus fruit on trees in low light and complex background based on MSER and HCA," *Trans. Chin. Soc. Agric. Eng.*, vol. 33, no. 19, pp. 196–201, Oct. 2017.
[23] E. Rachmawati, M. L. Khodra, and I. Supriana, "Fruit image segmentation by combining color and depth data," in *Proc. Int. Conf. Inf. Syst., Appl. Math.*, Jan. 2016, pp. 651–666.
[24] H. Liao, B. Wang, *Tutorial on Analytic Geometry*. 2nd ed. Beijing, China: Science Press, 2007, pp. 51–135.
[25] M. Camplani, T. Mantecon, and L. Salgado, "Depth-color fusion strategy for 3-D scene modeling with kinect," *IEEE Trans. on*, vol. 43, no. 6, pp. 1560–1571, Dec. 2013.
[26] H. Dang, J. Song, and Q. Guo, "A fruit size detecting and grading system based on image processing," in *Proc. Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, Nanjing, China, Aug. 2010, pp. 83–86.
[27] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. El Saddik, "Evaluating and improving the depth accuracy of Kinect for windows v2," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4275–4285, Aug. 2015.
[28] X. Wei, K. Jia, J. Lan, Y. Li, Y. Zeng, and C. Wang, "Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot," *Optik*, vol. 125, no. 19, pp. 5684–5689, 2014.
[29] J. Baeten, K. Donné, S. Boedrij, W. Beckers, and E. Claesen, "Autonomous fruit picking machine: A robotic apple harvester," in *Field and Service Robotics*, vol. 42. Berlin, Germany: Springer, 2008, pp. 531–539,
[30] T.-W. Chen, Y.-L. Chen, and S.-Y. Chien, "Fast image segmentation based on *K*-Means clustering with histograms in HSV color space," in *Proc. IEEE Workshop Multimedia Signal Process.*, Cairns, QLD, Australia, Oct. 2008, pp. 322–325.
[31] S. G. Kandi, "Automatic defect detection and grading of single-color fruits using HSV (hue, saturation, value) color space," *J. Life Sci.*, vol. 4, no. 7, pp. 39–45, Dec. 2010.

MIN HAN received the master's degree in control engineering from Beijing Forestry University, in 2018. His research interests include computer vision, machine learning, 3-D point cloud segmentation, image processing, and automatic picking.

JIANGMING KAN was born in 1976. He received the Ph.D. degree from Beijing Forestry University, Beijing, China, in 2009. He is currently a Professor and a Doctoral Tutor with Beijing Forestry University. His research interests include computer vision, pattern recognition, and intelligent systems.

YUTAN WANG was born in November 1974. He received the Ph.D. degree from the Beijing Forestry University of China, in 2014. He is currently a Professor with the School of Mechanical Engineering, Ningxia University of China. His research interests include artificial intelligence and image processing.