Research on tracking method of flower target

Yunqiang Chen¹, Qing Wang¹*, Hong Chen¹, Xiang Li¹ and Xiaoyu Song¹

¹ College of Information and Electrical Engineering, China Agricultural University, Beijing, 10083, China

*Corresponding author’s e-mail: wangqingait@sina.com

Abstract. Flowers are a common species in people’s lives, but people’s understanding of flowers is not comprehensive. The traditional UI display method lacks immersion and interactivity. This paper proposes a method based on the deep learning neural network flower recognition algorithm to recognize the flower target information and replace the traditional UI with the AR information to display the flower information. The tracking and positioning technology used in the floral AR display method is a very important underlying technology. This paper focuses on the tracking technology of flower target, and proposes a server-based flower recognition neural network algorithm recognition algorithm. After identifying the flower information and detection frame, it outputs the detection frame parameters of the flower target, and then combines the OpenCV KCF tracker pair. The method of performing AR display of the flower information has high tracking accuracy and stability.

1. Introduction

With the development of computer technology, augmented reality technology has gradually entered the public eye. The traditional graphic front-end display method can no longer meet the cultural and entertainment needs of the public, and the augmented reality information display method brings a fresh and interesting interactive experience. The current software such as “Xing Se Flower” has been popular among the masses, but its information display method for flowers still stays in the traditional UI interface. Compared with the traditional product display through image or video format, the use of AR technology can not only make the product display effect more realistic, but also add interactive modules, users can truly experience the product without leaving home. Certain features, which undoubtedly enhance the user experience of the product in marketing and display [1]. Target detection technology is the key point for the realization of augmented reality systems. Traditional methods cannot accurately identify objects in complex environments, and the pose estimation of objects is not accurate. With the development of target detection algorithms based on deep learning, models obtained through massive data training can quickly identify natural objects in the scene [2]. In order to solve the problem of recognition and tracking of moving targets in AR technology, Guo Yangyang [3] proposed a target recognition and tracking algorithm based on feature point matching result purification and iterative neighbor algorithm. In addition, in order to make AR tracking and other better operation on the mobile phone, Daniel Wagner [4] proposed three techniques for real-time 6 DOF natural feature tracking on the mobile phone, realizing an interactive frame rate of up to 30 Hz, which can be obtained from the current Textured planar targets on the phone for natural feature tracking.

The AR display method has been applied in many games [5], such as Nintendo’s “Pokémon Go” and Tencent’s "Catch the Demon". Compared with traditional 2D games and 3D games, AR games can
make players closer to the living environment, and integrate the game with the real environment to bring an immersive gaming experience.

2. Related work
The flower target tracking technology involves graphic coordinate systems, computer algorithms, and the like.

2.1. Introduction to four plane coordinate systems
Directions such as image processing and stereo vision often involve four coordinate systems: world coordinate system, camera coordinate system, image coordinate system, and pixel coordinate system.

The four coordinate systems involved in image processing:
- $O_w - X_wY_wZ_w$: world coordinate system, describing camera position, unit m
- $O_c - X_cY_cZ_c$: camera coordinate system, which optical center is the origin, unit m
- $o - xy$: image coordinate system, which optical center is the midpoint of the image, unit mm
Pixel coordinate system, the origin is the upper left corner of the image, unit pixel

P: one point in the world coordinate system is the true point in life

p: the imaging point of the point P in the image, the coordinates in the image coordinate system are \((x, y)\), and the coordinates in the pixel coordinate system are \((u, v)\)

\(f\): camera focal length, equal to the distance between \(o\) and \(O_c\), \(f = ||o - O_c||\)

2.2. Coordinate transformation of the tracking process

From the research in the previous section, we know that the implementation process of the virtual and real registration technology can be roughly divided into the following three coordinate transformation links.

2.2.1. Real space coordinate system → observation space coordinates. The coordinate transformation between the real space coordinate system and the observation space coordinate system is to determine the relative position and direction between the real scene and the user's head.

The conversion relationship between the real space coordinate system and the camera coordinate system can be expressed by the following formula:

\[
\begin{pmatrix}
X_c \\
Y_c \\
Z_c
\end{pmatrix} = \begin{pmatrix}
R & t \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
X_w \\
Y_w \\
Z_w
\end{pmatrix}
\]

(1)

The above formula reflects the Euclidean transformation relationship between the real space coordinate system and the camera coordinate system.

2.2.2. Observing the space coordinate system→2D imaging plane coordinate system. The coordinate transformation of the observation space coordinate system to the 2D imaging plane coordinate system is to obtain the coordinate conversion relationship of the camera to the 2D image plane.

According to the principle of linear model camera perspective projection, we know that for the pinhole camera model, it is assumed that the projection of \(P\) on the image is \(p(x, y)\), which is the intersection of \(O\) and \(P\) and the image plane. This relationship is called a central projection or a perspective projection and is derived from a similar proportional relationship:

\[
x = \frac{x_c f}{z_c} \\
y = \frac{y_c f}{z_c}
\]

(2)

Where \((X_c, Y_c, Z_c)\) is the coordinate of the spatial point \(P\) in the camera coordinate system, and \(f\) is the focal length of the camera. The above perspective projection relationship is expressed by homogeneous coordinates and a matrix as:

\[
\begin{pmatrix}
x \\
y \\
1
\end{pmatrix} = \begin{pmatrix}
f & 0 & 0 & 0 \\
0 & f & 0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
x_c \\
y_c \\
Z_c
\end{pmatrix}
\]

(3)

Since the image coordinates \((x, y)\) in the above equation are real numbers, and the image processing in the computer is in pixels, we should also convert the actual position of the dots in the image into pixel position. Let \((u, v)\) be the pixel coordinates of the image. The physical size of each pixel in the x-axis and y-axis directions is \(dx, dy\), then the relationship between \((x, y)\) and \((u, v)\) can be expressed as:

\[
u = \frac{x}{dx} + u_0 \\
v = \frac{x}{dy} + v_0
\]

(4)

Where \((u_0, v_0)\) represents the pixel center point of the image. For ease of expression, the above formula can be expressed in matrix form as:
\[
\begin{pmatrix}
\frac{1}{dx} & 0 & u_0 \\
0 & \frac{1}{dy} & v_0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x \\
y \\
1
\end{pmatrix}
\]
(5)

\[
\begin{pmatrix}
\frac{1}{dx} & 0 & -u_0d_x \\
0 & \frac{1}{dy} & -v_0d_y \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
u \\
v \\
1
\end{pmatrix}
\]
(6)

Substituting equation (5) into equation (3), we can get the conversion relationship between camera 3D coordinates and image plane 2D coordinates, as follows:

\[
Z_c
\begin{pmatrix}
u \\
v \\
1
\end{pmatrix}
= \begin{pmatrix}
\frac{1}{dx} & 0 & u_0 \\
0 & \frac{1}{dy} & v_0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
f & 0 & 0 & 0 \\
f & 0 & 0 & 0 \\
0 & 1 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
X_c \\
Y_c \\
Z_c \\
1
\end{pmatrix}
\]
\[
= \begin{pmatrix}
a_x & 0 & u_0 & 0 \\
0 & a_y & v_0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
X_c \\
Y_c \\
Z_c \\
1
\end{pmatrix}
= \begin{pmatrix}
a_x & 0 & u_0 & 0 \\
0 & a_y & v_0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
X_v \\
Y_v \\
Z_v \\
1
\end{pmatrix}
\]
(7)

Where \(a_x = f * dx\), \(a_y = f * dy\), since \(a_x\), \(a_y\), \(u_0\), \(v_0\) are only related to the internal structure of the camera, these parameters are called internal parameters of the camera.

2.2.3. Virtual object coordinate system \(\rightarrow\) real space coordinate system. The geometric description of the three-dimensional virtual object in \(O_v - X_vY_vZ_v\) is transformed into the real space coordinate system. The transformation relationship of the geometric description in \(O_w - X_wY_wZ_w\) can be expressed by the following formula:

\[
\begin{pmatrix}
X_w \\
Y_w \\
Z_w \\
1
\end{pmatrix}
= A
\begin{pmatrix}
X_v \\
Y_v \\
Z_v \\
1
\end{pmatrix}
\]
(8)

In the above formula, \(A\) is a transformation matrix known between two coordinate systems. Through the above three coordinate links, we can theoretically find the position that the virtual object should appear in the image plane. Substituting equation (7) into equation (8) yields the following equation:

\[
Z_c
\begin{pmatrix}
u \\
v \\
1
\end{pmatrix}
= A
\begin{pmatrix}
X_v \\
Y_v \\
Z_v \\
1
\end{pmatrix}
= N
\begin{pmatrix}
X_v \\
Y_v \\
Z_v \\
1
\end{pmatrix}
\]
(9)

It can be seen from the above formula that we only need to obtain the coordinate transformation matrix \(N\), and we can map each point of the virtual object one by one to the image plane to realize the virtual real superposition function.

2.3. Introduction to OpenCV KCF Algorithm

KCF is an authentication tracking method. These methods generally train a target detector during the tracking process, use the target detector to detect whether the predicted position of the next frame is the target, and then use the new detection result to update the training set. KCF's main contribution include: 1) The dot multiplication greatly reduces the amount of calculation, improves the operation speed, and makes the algorithm meet the real-time requirements. 2) The ridge regression of the linear
space is mapped to the nonlinear space by the kernel function. 3) A way to integrate multi-channel data into the algorithm is presented.

3. Flower target tracking

Starting with OpenCV 3, the target detector has evolved rapidly. Here are the advantages and disadvantages of these types of trackers:

- MIL Tracker: More accurate than the previous tracker, but the failure rate is higher.
- KCF Tracker: Faster than BOOSTING and MIL, but not well in the case of occlusion.
- CSRT Tracker: Slightly more accurate than KCF, but not as fast as the latter.
- Median Flow Tracker: Work well in error reporting, but for fast-moving or fast-moving objects, the model will fail.
- MOSSE Tracker: It's fast, but not as good as the accuracy of CSRT and KCF.

Comparing the advantages and disadvantages of the above OpenCV tracker, combined with the characteristics of the flower target, the selected tracker is the KCF tracker.

Tracking floral targets using KCF is a feature point based on tracking method, while general steps for feature point based on target tracking include:

1. Detecting feature points of the current frame.
2. The position of the current frame feature point in the next frame is estimated by the gray level comparison of the current frame and the next frame.
3. Filter the feature points with the same position, and the remaining points are the targets.

The feature point-based target tracking algorithm is related to two steps (1) and (2). The following is an improved Harris corner to extract feature points, and the Lucas-Kanade optical flow method is used to achieve target tracking.

The optical flow method [6] associates the two-dimensional velocity field with the gray level and introduces the optical flow constraint equation to obtain the basic algorithm for optical flow calculation. The optical flow is the "instantaneous velocity" of the pixel motion of a spatially moving object on an observed imaging plane.

The optical flow method for flower target tracking steps is:

1. The Unity3D client obtains the first frame image with a pixel size of 400*400.
2. Upload (1) the obtained image to the server, input to the flower target detection neural network model, and output the position coordinates of the flower in the single frame picture.
3. Initialize the coordinates of the flower detection frame obtained in (2) to the OpenCV KCF tracker and start updating the tracker to track the changing target position.
4. The KCF tracker processes a sequence of consecutive video frames containing floral objects.
5. For each video sequence, the KCF tracker is used to detect possible foreground targets.
6. If a foreground target appears in a frame, find its representative key feature points.
7. For any two adjacent video frames that follow, the optimal position of the key feature points appearing in the previous frame in the current frame is found, thereby obtaining the position coordinates of the foreground target in the current frame.
8. By doing this iteratively, the tracking of the flower target can be achieved.

4. Experimental result

The experimental operating environment is Windows10, the processor frequency is 3.60GHz, and the lens resolution is set to 1280*720. The client platform is Unity3D, the programming language is C#, the flower target detection neural network model runs on Tencent cloud server, and the OpenCV KCF tracker runs on the client. The picture recognition and picture tracking of the flower target in the experiment are carried out under two threads. The process of picture recognition is about 50ms. If the picture tracking is run under a single thread, the tracking time of each frame is about 20ms. The experimental results are shown in Figures 6, 7, 8, 9, 10, 11.
The KCF model was used to track and locate the flower target. The tracking effect experiments were carried out on the front, left side, right side, upper side and lower side of the flower image, respectively, which showed good accuracy and stability.

5. Conclusion

Based on the known deep learning neural network flower target detection model as the flower target recognition detection, this paper proposes a method based on the former recognition output detection frame coordinate result after a series of coordinate system conversion to initialize the OpenCV KCF tracker. The floral target position coordinates of the frame, the KCF tracker continuously updates the tracked flower target position coordinates through a series of tracking algorithms. The experimental results show that the method can quickly track the identified flower targets and update the coordinates of the flower targets, which is of great significance to the AR registration rendering of subsequent flower targets.

References

[1] Li J.W, (2018) Design and Implementation of Virtual Vehicle Display System Based on Augmented Reality Technology, South China University of Technology. Page 65.
[2] Zhang S.B, (2018) Design and Implementation of Augmented Reality System Based on SSD Object Tracking Algorithm, Northwest University. Page 61.
[3] Guo Y.Y, Sun H, (2018) Target Recognition and Tracking Algorithm Based on Monocular AR Environment. Computer Technology and Development, 28(10): Pages 8-12.
[4] Wagner, D., et al., (2010) Real-time detection and tracking for augmented reality on mobile phones. IEEE Transactions on Visualization and Computer Graphics, 16(3): p. 355-368.
[5] OurPlay. (2018) In addition to Pokemon Go, which AR games are worth playing? Https://baijiahao.baidu.com/s?id=1620280911233872522&wfr=spider&for=pc
[6] Little brother th. (2017) The use of optical flow tracker in OpenCV learning Https://blog.csdn.net/ZHANG781068447/article/details/78565078