Application of Physical Interactive Mixed Reality System Based on MLAT in the Field of Stage Performance

Yanxiang Zhang(✉), Pengfei Ma, Ali Raja Gulfraz, Li Kong, and Xuelian Sun

Department of Communication of Science and Technology, University of Science and Technology of China, Hefei, Anhui, China
petrel@ustc.edu.cn

Abstract. At present, in some real-time occasion like as stage performances, mixed reality effects are often generated by a mechanical means, or only between physical objects and virtual objects. This will lead to many uncontrollable shortcomings. in order to overcome the above disadvantage, the authors developed a mixed reality system for physical interaction. The experiment unfolds between the experiencer and the controllable Four-axis unmanned aerial vehicle with some performance props. The first step of the study is to build the stereoscopic scene through the camera calibration, the target coordinates in image are obtained by infrared sensors deployed on the experimenter and the aircraft respectively by Camshift algorithm and it can be transformed into the spatial coordinates by projection transformation. And the second step is based on different preset working modes and the coordinate relationship expressions between them that have been configured, with the action of the experimenter, Ant Colony Algorithm is used to move the Four-axis UAV to a certain position, thereby realizing this kind of precisely controllable interaction between entities. The final experiment in this paper has proven that our work provided a great mixed reality effect for occasions such as stage performances.

Keywords: Infrared sensor · Camera calibration · Camshift · Route plan · Four-axis UAV · OpenCV

1 Introduction

In recent years, with the continuous development of technologies such as digitization, visualization and mixed reality, people have put forward higher requirements for the experience of real-time performance on the stage. In the traditional stage performance field that combines mixed reality technology, the audience's experience and satisfaction are mainly from the interaction between actors and virtual props [1]. But so far the technology of superimposing virtual elements on the exact position of a real scene by a computer is still immature and requires manual assistance to achieve. In practical applications, taking the stage performances of ancient Chinese mythology as an example, these mythological backgrounds require some special effects, which often need the interaction of real actors and props to produce. In the current means of producing these special effects, one of the means is to let the actors fly by using

L. T. De Paolis and P. Bourdot (Eds.): AVR 2019, LNCS 11614, pp. 3–12, 2019.
https://doi.org/10.1007/978-3-030-25999-0_1

mechanical techniques, and another current mainstream method uses actors to interact with virtual images in the background of the screen [2], the above means often have the characteristics of low dynamic response resolution, insufficient degree of freedom and process-uncontrollability. In view of the improvement of the above shortcomings, authors developed an entity mixed reality interactive system to apply to occasions such as stage performances.

In this system, experiment was unfolded between the experiencer and the controllable Four-axis unmanned aerial vehicle, which some physical items can be overlaid or hung underneath. The system mainly aims to accomplish such a function: Under different mode settings, the drone with the physical props can move according to a certain path as the entity moves. Of course, the entity here is a hand but is not limited to this part. In order to achieve this function, the system is mainly divided into the following sections:

I. The infrared binocular camera is used to detect the infrared light source deployed on the experimenter's hand and the drone. The detection algorithm uses the Camshift algorithm to obtain the 2D coordinates of the two moving targets on one frame of the video stream in real time.
II. Calibrating the binocular camera to obtain the stereo space where the moving target is located, and converting the 2D coordinates (x, y) into real-time 3D coordinates (x, y, z).
III. Through the movement of the experimenter's hand, the real-time $P_{hand} = (x_{hand}, y_{hand}, z_{hand})$ coordinates are obtained. According to the preset flight path relationship $P_{UAV} = f(P_{hand})$ between the two targets, we can combined with the flight control system command and Ant Colony Algorithm and adjusted the drone to $P_{UAV} = (x_{UAV}, y_{UAV}, z_{UAV})$ in real time.

Here, we will give the overall design framework of the system and elaborate on the theoretical details and engineering implementation details. Figure 1 describes it.

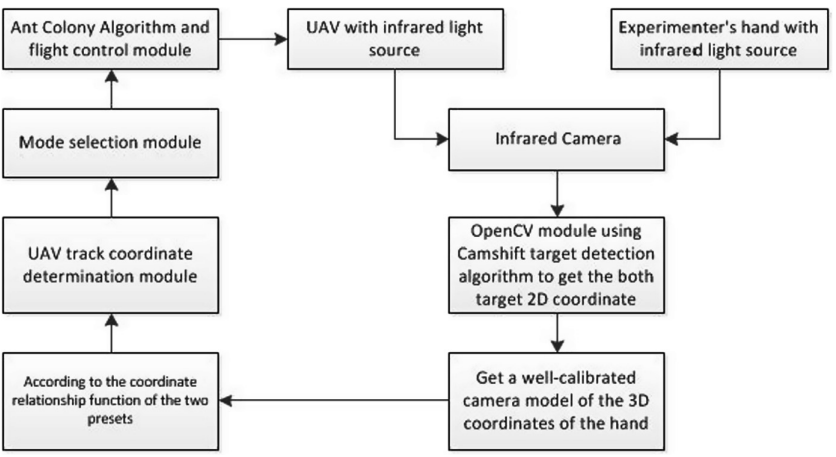

Fig. 1. System overall flow chart.

2 System Design

2.1 Camshift Moving Target Detection Algorithm

In our system, the 2D coordinate tracking of two infrared moving target points (hand and drone) uses the Camshift algorithm. As we all know, Camshift uses the target's color histogram model to convert the image into a color probability distribution map. It can initialize the size and position of a search window, and adaptively adjust the position and size of the search window based on the results obtained in the previous frame, thereby locate the central location of the target in the current image [3]. The system realizes the tracking of 2D coordinates mainly adopts the following three processes.

I. (1) To avoid image sensitivity to light, we convert the image from RGB space to HSV space.
   (2) By calculating the histogram of the H component, the probability or number of pixels representing the occurrence of different H component values is found in the histogram, which can get the color probability look-up table.
   (3) A color probability distribution map is obtained by replacing the value of each pixel in the image with the probability pair in which the color appears. Actually this is a back projection process and color probability distribution map is a grayscale image.

II. The second process of the Camshift algorithm uses meanshift as the kernel. The meanshift algorithm is a non-parametric method for density function gradient estimation. It detects the target by iteratively finding and optimizing the extreme value of the probability distribution. We use the following flow chart to represent this process.

III. Extending the meanshift algorithm to a continuous image sequence is achieved by the camshift algorithm [4]. It performs a meanshift operation on all frames of the images, and takes the result of the previous frame. That is, the size and center of the search window becomes the initial value of the search window of the next frame of the meanshift algorithm. With this iteration, we can track 2D coordinates of the experimenter's hand and drone on the image. Its main following process is based on the integration of I and II.

In the engineering implementation of the above algorithm, we developed a MFC-based PC software using the Camshift function in OpenCV as a kernel to display the 2D coordinates of the two moving target in real time.

2.2 3D Reconstruction

In order to obtain the 3D coordinates of the hand and drone in the world coordinate system, we carried out a 3D reconstruction experiment based on camera calibration. Before understanding 3D coordinates, we need to understand the four coordinate systems and the three-dimensional geometric relationship between them. First, we introduce 3D reconstruction based on binocular camera.

I. Regarding the calibration of the camera, we should first understand the four coordinate systems.

(1) Pixel coordinate system. The Cartesian coordinate system u-v is defined on the image, and the coordinates (u, v) of each pixel are the number of columns and the number of rows of the pixel in the array. Therefore, (u, v) is the coordinate of image coordinate system in pixels, which is also the (x, y) value obtained by the Camshift algorithm in our previous system module.

(2) Retinal coordinate system. Since the image coordinate system only indicates the number of columns and rows of pixels in the digital image [5], and the physical position of the pixel in the image is not represented by physical units, it is necessary to establish an retinal coordinate system x-y expressed in physical units (for example, centimeters). we use (x, y) to represent the coordinates of the retinal coordinate system measured in physical units. In the x-y coordinate system, the origin O_1 is defined at the intersection of the camera's optical axis and the image plane, and becomes the principal point of the image [6]. This point is generally located at the center of the image, but there may be some deviation due to camera production. O_1 becomes (u_0, v_0) in the coordinate system, and the physical size of each pixel in the x-axis and y-axis directions is dx, dy. The relationship between the two coordinate systems is as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & s' & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

where s' represents the skew factor because the camera retinal coordinate system axes are not orthogonal to each other.

(3) World coordinate system. The relationship between the camera coordinate system and the world coordinate system can be described using the rotation matrix R and the translation vector t. Thus, the homogeneous coordinates of the point P in the space in the world coordinate system and the camera coordinate system are $(X_w, Y_w, Z_w, 1)^T$ and $(X_c, Y_c, Z_c, 1)^T$, respectively, and the following relationship exists:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

where R is a $3x3$ orthogonal unit matrix, t is a 3-dimensional translation vector, and $0 = (0, 0, 0)^T$, M_1 is the contact matrix between two coordinate systems. In our opinion, there are twelve unknown parameters in the M_1 contact matrix to be calibrated. Our method is to use the twelve points randomly selected by space, and find the coordinates of these twelve points according to the world

coordinates and camera coordinates, then it forms twelve equations, so the M_1 matrix can be uniquely determined.

(4) Camera linear model. Perspective projection is the most commonly used imaging model and can be approximated by a pinhole imaging model [7]. It is characterized in that all light from the scene passes through a projection center, which corresponds to the center of the lens. A line passing through the center of the projection and perpendicular to the plane of the image is called the projection axis or the optical axis. As shown in Fig. 2, x_1, y_1 and z_1 are fixed-angle coordinate systems fixed on the camera. Following the right-hand rule, the X_c axis and the Y_c axis are parallel to the coordinate axes x_1 and y_1 of the image plane, and the distance OO_1 between the planes of the X_c and Y_c and the image plane is the camera focal length f. In the actual camera, the image plane is located at the distance f from the center of the projection, and the projected image is inverted. To avoid image inversion, it is assumed that there is a virtual imaging x', y', z' plane in front of the projection center. The projection position (x, y) of $P(X_c, Y_c, Z_c)$ on the image plane can be obtained by calculating the intersection of the line of sight of point $P(X_c, Y_c, Z_c)$ and the virtual imaging plane.

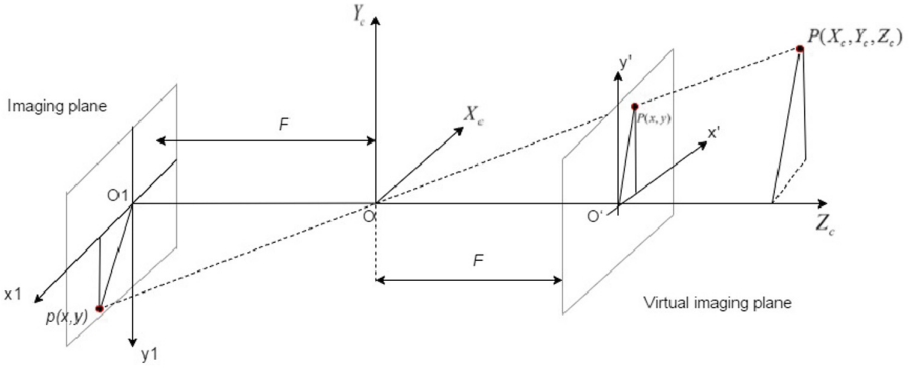

Fig. 2. Camera model.

The relationship between the camera coordinate system and the retinal coordinate system is:

$$x = \frac{fX_c}{Z_c}, y = \frac{fY_c}{Z_c}$$

where (x, y) is the coordinate of point P in the retinal coordinate system, and $P(X_c, Y_c, Z_c)$ is the coordinate of the space point P in the camera coordinate system, which is represented by the subordinate coordinate matrix:

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

By combining the above equations, we can get the relationship between the image coordinate system and the world coordinate system:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & s' & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = K[R \quad t]\tilde{X} = P\tilde{X}$$

where $\alpha_u = \frac{f}{dx}$, $\alpha_v = \frac{f}{dy}$, $s = s'f$. $[R \quad t]$ is completely determined by the orientation of the camera relative to the world coordinate system, so it is called the camera external parameter matrix, which consists of the rotation matrix and the translation vector. K is only related to the internal structure of the camera, so it is called the camera internal parameter matrix. (u_0, v_0) is the coordinates of the main point, α_u, α_v are the scale factors on the u and v axes of the image, respectively, s is the parameter describing the degree of tilt of the two image coordinate axes [8]. P is the 3×4 matrix called the projection matrix, that is, conversion matrix of world coordinate system relative to image coordinate system. It can be seen that if the internal and external parameters of the camera are known, the projection matrix P can be obtained. For any spatial point, if the three-dimensional world coordinates (X_w, Y_w, Z_w) are already known, the position (u, v) at the image coordinate point can be obtained. However, if we know the image coordinates (u, v) at a certain point in the space even if the projection matrix is known, its spatial coordinates are not uniquely determined. In our system, it is mainly determined by using a binocular camera to form stereoscopic vision and depth information to get the position of any point in the world coordinate.

II. 3D reconstruction based on the binocular camera. As we know, when people's eyes are observing objects, the brain will naturally produce near and deep consciousness of the object. The effect of generating this consciousness is called stereo vision. By using a binocular camera to observe the same target from different angles, two images of the target can be acquired at the same time. And the three-dimensional information is restored by the relative parallax of the target in the imaging, thereby realizing the stereoscopic positioning effect.

As shown in Fig. 3, for any point P in space, two cameras C_1 and C_2 are used to observe point P at the same time.O_1 and O_2 are the optical centers of the two cameras respectively. P_1, P_2 are the imaging pixels of P in the imaging plane of two cameras. It

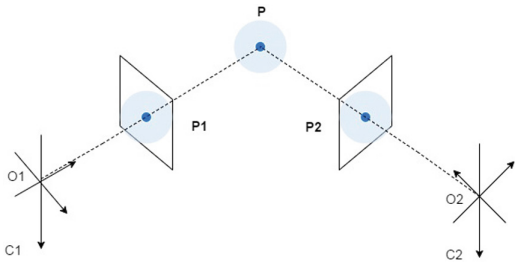

Fig. 3. Binocular vision imaging principle.

can be known that the straight line O_1P_1 and the straight line O_2P_2 intersect at the point P, so the point P is unique and the spatial position information is determined.

In this model, the three-dimensional coordinate calculation of the spatial point P can be solved by the least squares method according to the projection transformation matrix.

Assuming that in the world coordinate system, the image coordinates of the spatial point $P(x, y, z)$ on the imaging planes of the two cameras are $P_1(u_1, v_1)$ and $P_2(u_2, v_2)$. According to the camera pinhole imaging model, we can get:

$$Z_{c1}\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = M_1\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, Z_{c2}\begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = M_2\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix},$$

$$M_1 = \begin{bmatrix} m_{111} & m_{112} & m_{113} & m_{114} \\ m_{121} & m_{122} & m_{123} & m_{124} \\ m_{131} & m_{132} & m_{133} & m_{134} \end{bmatrix}, M_2 = \begin{bmatrix} m_{211} & m_{212} & m_{213} & m_{214} \\ m_{221} & m_{222} & m_{223} & m_{224} \\ m_{231} & m_{232} & m_{233} & m_{234} \end{bmatrix}$$

where Z_{c1} and Z_{c2} are the Z coordinates of the P points in the left and right camera coordinate systems, and M_1 and M_2 are the projection matrices of the left and right cameras. The premise of these two formulas is that we must obtain the pixel coordinates (u_1, v_1) and (u_2, v_2) on the left and right images of P point in advance. We combine the two equations above and then eliminate Z_{c1} and Z_{c2}. Then we will get:

$$AP = b$$

where:

$$A = \begin{bmatrix} m_{131} - m_{111} & m_{132} - m_{112} & m_{133} - m_{113} \\ m_{131} - m_{121} & m_{132} - m_{122} & m_{133} - m_{123} \\ m_{231} - m_{211} & m_{232} - m_{212} & m_{233} - m_{213} \\ m_{231} - m_{221} & m_{232} - m_{222} & m_{233} - m_{223} \end{bmatrix}, P = [x \quad y \quad z]^T, b = \begin{bmatrix} m_{114} - u_1 m_{134} \\ m_{124} - v_1 m_{134} \\ m_{214} - u_2 m_{234} \\ m_{224} - v_2 m_{234} \end{bmatrix}$$

According to the least squares method, the three-dimensional coordinates of the spatial point P under the world coordinate system can be obtained as:

$$P = (A^T A)^{-1} A^T b$$

Therefore, we firstly use Camshift to find the pixel coordinates (u_{hand}, v_{hand}) of the experimenter's hand on the image through the above series of algorithms, then 3D reconstruction is performed by binocular camera vision to obtain the 3D coordinates of the hand in the world coordinate system.

2.3 Ant Colony Algorithm for Path Planning

In view of the fact that the world coordinates of the space field points are already available, the next problem to be solved is to use the coordinates of the hand so that the drone can move with it. We extracted it as a path planning problem for drones. System uses the Ant Colony Algorithm to adjust the two moving targets for precise motion through the preset interaction path between the hand and the drone.

The algorithm in this paper is to set several points on the preset path, which are randomly generated by the movement of the hand, and the number is not fixed, so this is a typical TSP problem. In our system, the central controls the drone to fly in accordance with the path of the random point coordinates generated by the hand. When the UAV position is off the route or the target point is changed, the system will regenerate the path based on the changed UAV dynamic and static information using the Ant Colony algorithm. When the flight path of the drone changes, the system will respond quickly. According to the newly generated track of the drone system, the drone will be controlled to fly. System will convert the control command signal into a PWM signal to reach the drone and realize the purpose of attitude control. The implementation of the entire aircraft control system is shown in Fig. 4.

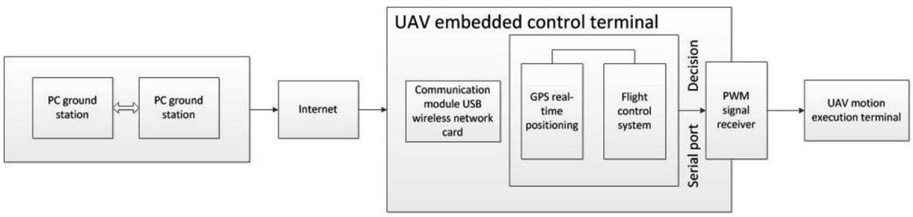

Fig. 4. UAV path planning system based on ant colony algorithm.

3 Field Experiment

The field experiment of the system is mainly divided into three parts. The first parts is the calibration of the binocular camera, then the second part are construction and deployment of the UAV path planning system and sensor components, which

implement the two main modes: including that drone props flying around space and interact with actors' actions, and the third part is the physical interaction between the experimenter and the drone props. Here are some of our system experiments pictures.

First, the deployment of the UAV terminal control system and sensor system are designed by the authors. This work is the basis for the debugging of experimental hardware systems.

The Fig. 5 shows photos of an experiencer interacting with a drone with a balloon prop. The system detects the coordinates of the change of the hands of the experiencer and adjusts the drone to its corresponding coordinates in real time. The effect is that the experiencer can push the balloon to fly and realize the physical interaction between both.

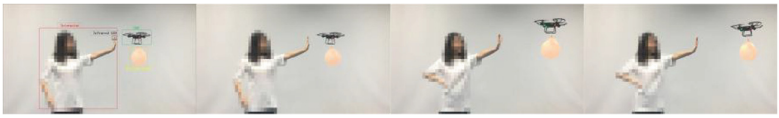

Fig. 5. Experiments in which researchers interact with physical items.

After joint debugging, it is confirmed that the system can achieve dynamic interaction between entities within a certain range, but the time response still has certain limitations. The authors are prepared to solve this problem with a faster algorithm in the next work.

4 Result

This paper develops a mixed reality system for scene interaction between actors and solid props in the stage performance field. The implementation of the experiment combines multiple techniques of multi-point positioning, computer vision, drone control, and path planning. Through the previous technical details and field experiments, authors finally realized the scene of the drone props flying around the stage space and interacting with the actors according to the set mode. These implementations have greatly overcome some of the limitations of special effects in traditional stage performances and enhance the immersion of the audience. It has great potential for future business deployment.

References

1. Chatman, S.B.: Story and Discourse: Narrative Structure in Fiction and Film. Cornell University Press, Ithaca (1980)
2. Zhang, Y., Ma, P.F., Zhu, Z.Q.: Integrating performer into a real-time augmented reality performance spatially by using a multi-sensory prop. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, p. 66. ACM (2017)

3. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. Int. J. Comput. Vis. **92**(1), 1–31 (2011)
4. Allen, J.G., Xu, R.Y.D., Jin, J.S.: Object tracking using camshift algorithm and multiple quantized feature spaces. In: Proceedings of the Pan-Sydney area workshop on Visual information processing, pp. 3–7. Australian Computer Society, Inc. (2004)
5. Yimin, L., Naiguang, L., Xiaoping, L., Peng, S.: A novel approach to sub-pixel corner detection of the grid in camera calibration. In: 2010 International Conference on Computer Application and System Modeling (ICCASM), vol. 5, pp. V5–18. IEEE (2010)
6. Li, J., Wan, H., Zhang, H., Tian, M.: Current applications of molecular imaging and luminescence-based techniques in traditional Chinese medicine. J. Ethnopharmacol. **137**(1), 16–26 (2011)
7. Wen, Z., Cai, Z.: Global path planning approach based on ant colony optimization algorithm. J. Central South Univ. Technol. **13**(6), 707–712 (2006)
8. Bouguet, J.-Y.: Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. Intel Corporation **5**(1–10), 4 (2001)