

Real-time Projection Method for Augmented Reality Assisted Assembly

Guo Hong-jie¹, Du Fu-zhou¹

¹School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China

Abstract. Recently, the augmented reality technology has become a useful tool for assembly guidance. The projectors have been always used as virtual image output devices. In most situations, real-time and dynamic images projection is essential due to that the components to be assembled are randomly placed and movable. However, the cameras and the projectors are placed in different relative positions, making it difficult to project real time images when we are using augmented reality for assembly. A novel method based on the system of binocular cameras and projector was proposed here to overcome the limitation. We established a method to get the relations of coordinate transform among camera coordinate system, projector coordinate system and world coordinate system based on real-time internal parameter matrix of the projector that we derived. Obtaining the pose information of the cameras without any designed markers in real world was also realized, which is the key technology for the camera-projector assembly visualization system. An assembly experiment of cable laying was conducted and the results showed that using the proposed method the real-time projection for augmented reality assisted assembly was realized.

1 Introduction

The assembly processes of aerospace products are characterized by complexity, diversity and long assembly cycles. In the traditional assembly process, engineers and technicians mainly rely on complicated 2D drawings and technical manuals to complete the assembly work. This not only has the problem of relatively low assembly efficiency, but also makes the engineering quality easily affected by operational errors. With the development of the digital technology, computer technology is gradually used to guide assembly referring to three-dimensional models of product components. Since the augmented reality (AR) was proposed in the 1990s, it had been applied to Boeing's wiring system. AR technology is a technology that superimposes virtual digital model information onto a real environment to enhance visual effects. In most cases, the AR-assisted assembly includes the following processes:

1. Using the camera to capture the image of the real scene.
 2. Obtaining the camera's three-dimensional pose information in the real world.
 3. Superimposing the virtual model onto the pictures of the real scene according to the pose information above.
 4. Getting a virtual integrated image or video and display it on a helmet display or an ordinary display.
- All processes are shown in Figure. 1.

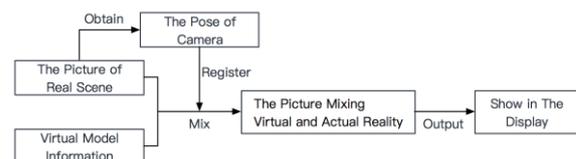


Figure 1. Basic flowchart of AR-assisted assembly

Reiners D et al. used AR technology to develop a real-time three-dimensional system to guide workers to install a door lock into a car door [1]. Yin Xuyue et al. have developed an integrated training system that included the operation guidance of the assembling and the function of testing and recording the key components to meet the demands of aerospace product assembling and training guidance in the production process [2]. The VVT company proposed a helmet display for AR-assisted assembly operation [3]. The operator could use the helmet to see a picture where the virtual parts and the prompt information are fused. Yao Yuan provided an AR application using a projector as an output device [4]. In this case, a merged image or video was projected onto a real scene through a projection device, so as to achieve the purpose of enhancing the visual effect. Chen Xiang-hui et al. carried out a method of extracting feature points from reference image and real scene image. The method was proposed to achieve the pose of the camera for the purpose of real-time projection [5].

For the assembly process of large components such as the installation of the fasteners of the aircraft, the wiring of electrical devices, etc., assembly training using AR technology cannot serve as a clear guidance. On the other hand, the position of the camera in the helmet

* Corresponding author: du_fuzhou@163.com

display did not coincide with the view of the human eye. Therefore, there was a certain deviation between the virtual scene and the scene seen by the human. In addition, the head-mounted display may cause dizziness, dazzling, nausea and other symptoms when the user's head is unbalanced. So it is not suitable for long-time wearing [6]. The use of a projector as an output device [4] can not only assist the assembly work intuitively, but also avoid the discomfort that may be caused by wearing a head-mounted display. However, the components to be assembled and the projection device were required to be fixed after the system configuration is completed using this method, which was an obvious limitation. In Chen Xiang-hui's research, the pose of the camera was impossible to obtain for lacking reference image matching the real scene [5]. To solve the problems above, this paper proposes an AR projector method based on binocular vision, which not only solves the problem of view deviation between the camera and the projection device, but also overcomes the limitation that the position of the assembly scene cannot be changed without the use of references. It makes AR-assisted assembly more practical.

2 Method

It can be seen from Figure. 1 that the virtual and real mixed image is obtained according to the pose information of the camera in the real-world coordinate system. In other words, the image is from the camera's view. However, because the position of the camera and the projector does not coincide, which means that the field of view between the camera and the projector is different, after being combined with the actual image, the image needs to be converted into an image from the field of view of the projector in order to accurately achieve visual enhancement. Meanwhile, taking into account the practical problems of the assembly, the conversion of the field of view between the camera and the projector needs to adapt to the change of the assembly environment. The conversion matrix between the camera and the projector will change as the assembly environment changes. This paper proposes the following AR-assisted assembly scheme using a projector as a display device. All the processes are shown in Figure. 2.

Common Process of Augmented Reality

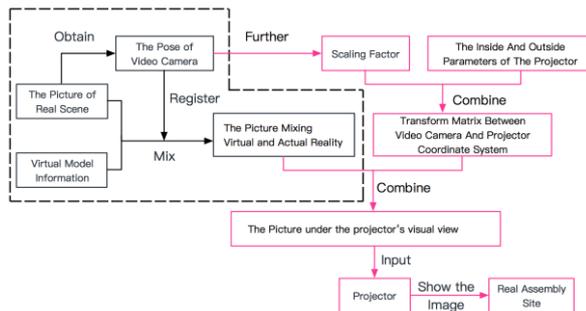


Figure 2. Flow chart of AR-assisted assembly using projector as display device

The solution in this paper refers to the general AR-assisted assembly process, which can obtain the camera

position. The camera depth information is also obtained. We combine the internal and external parameter matrix of the projector and obtain a dynamic field of view conversion matrix between the camera and the projector. Using the matrix, we obtain the image or video under the field of view of the projector, and finally display image or video. Thus, the key process is to obtain the inside and outside parameter matrix of the projector and the depth information of the camera, and finally achieve the purpose of calibrating the camera and the projector.

2.1 Vision-based camera and projector pose conversion method

In order to quantitatively describe the imaging process of the camera, a coordinate system consisting of a real-world coordinate system $O_w (X_w, Y_w, Z_w)$, a camera coordinate system $O_c (X_c, Y_c, Z_c)$, an image physical coordinate system $O (x, y)$, and an image pixel coordinate system $O_p (u, v)$ are defined, as shown in Figure. 3. The real-world coordinate system is an arbitrarily-defined three-dimensional space coordinate system. The camera coordinate system is a three-dimensional coordinate system based on the camera's optical center. The physical coordinate system of the image is a two-dimensional plane coordinate system whose origin is the intersection of the optical axis and the image plane. The image pixel coordinate system is a computer frame memory coordinate system in pixels with the origin in the upper left corner of the image.

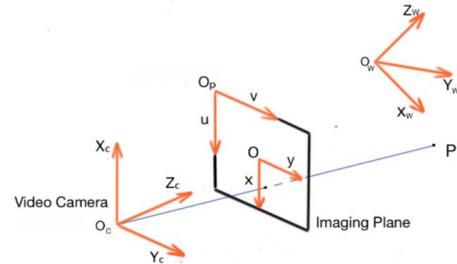


Figure 3. Schematic diagram of each coordinate system of camera imaging model

Let the coordinates of point P in space be $\tilde{P}_w(X_w, Y_w, Z_w, 1)^T$ in the space coordinate system and $\tilde{P}_c(X_c, Y_c, Z_c, 1)^T$ in camera coordinate system, then the conversion relationship between space coordinate system and camera coordinate system is:

$$\tilde{P}_c = [R \ T] \tilde{P}_w \quad (1)$$

R and T are the rotation and translation vectors of the world coordinate system relative to the image coordinate system.

Let the coordinates of the point under the corresponding image pixel coordinate system be $\tilde{p}(u, v, 1)^T$, then the conversion relationship between the camera coordinate system and the image pixel coordinate system can be expressed as:

$$\tilde{s}\tilde{p} = A[R \ T] \tilde{P}_w = A\tilde{P}_c \quad (2)$$

In the formula (2), s is the scaling factor, which represents the depth information of the pixel; A is the camera's internal parameter matrix.

The physical focal length of the camera lens is defined as f . The lengths of the unit pixels in the u -axis and v -axis of the image pixel coordinate system are respectively s_u and s_v . The lens focal lengths in the u -axis and v -axis are described by the number of pixels, whose formula are $f_u=f/s_u, f_v=f/s_v$; In fact, the camera's optical center is not on the optical axis, so it is necessary to introduce two new parameters c_u and c_v to describe the shift of the optical center. The axis of the imaging plane of the camera may incline, we use γ to represent the u axis and v axis deviation. Thus, the camera's internal parameter matrix can be expressed as:

$$A = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

In addition, the camera lens will cause radial distortion of the image, and the camera assembly process will cause tangentially distorted. So, the problem of lens distortion is also taken into consideration during camera calibration.

For radial distortion, the distortion of the center of the image is 0, and the closer the pixel point is to the edge of the image, the more severe the distortion will be. However, in practice, this distortion is not large, so it can be quantitatively described by the first few terms of the Taylor series expansion around the location. Then, the physical coordinates of a point on the imaging plane corrected according to radial distortion can be expressed as:

$$x' = x(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad (4)$$

$$y' = y(1 + k_1r^2 + k_2r^4 + k_3r^6) \quad (5)$$

Tangential distortion is caused by the fact that the lens itself is not parallel to the imaging plane. Using parameters p_1 and p_2 to describe the tangential distortion, the coordinates of the corrected point can be expressed as:

$$x'' = x + [2p_1y + p_2(r^2 + 2x^2)] \quad (6)$$

$$y'' = y + [p_1(r^2 + 2y^2) + 2p_2x] \quad (7)$$

In summary, the lens distortion considers five distortion factors k_1, k_2, k_3, p_1 and p_2 . These five distortion parameters also belong to the camera's internal parameters.

According to the reversibility of the optical path, the projector can be regarded as a reverse camera [7]. Then, a three-dimensional coordinate system of the projector can be established with the imaginary optical center O_p of the projector as an origin, and an image pixel coordinate system can be established on the virtual imaging plane. The problem of field of view conversion between the camera and the projector can be regarded as the problem of obtaining the pixel coordinate transformation matrix between the camera and the projector. The coordinate system of the camera and the projector is shown in Figure. 4.

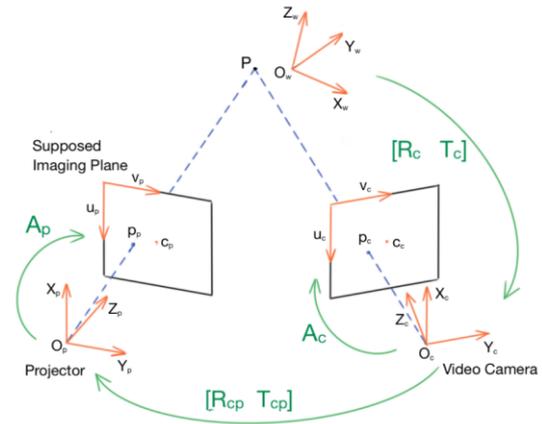


Figure 4. Schematic diagram of pixel and pixel coordinate transformation between camera and projector

Let the space point P be known as the corresponding point in the pixel coordinate system of the camera imaging plane, which is $\tilde{p}_c(u_c, v_c)$. According to the formula (2), the coordinate of the point in the camera coordinate system can be expressed as:

$$\tilde{P}_c = \frac{s}{A_c} \tilde{p}_c \quad (8)$$

Where A_c is the camera internal parameter, s is the scaling factor, which contains the depth-related proportion information.

According to the camera model, the transformation between the camera coordinate system and the projector coordinate system also has the transformation matrix R_{cp} and the translation matrix T_{cp} , then the coordinates of the point under the projector coordinate system can be expressed as:

$$\tilde{P}_p = [R_{cp} \quad T_{cp}] \tilde{P}_c = [R_{cp} \quad T_{cp}] \frac{s}{A_c} \tilde{p}_c \quad (9)$$

Assuming that the projector assumes that the internal reference matrix A_p is known, the point on the pixel coordinate of the projector can be expressed as:

$$\tilde{p}_p = A_p \tilde{P}_p = A_p [R_{cp} \quad T_{cp}] \frac{s}{A_c} \tilde{p}_c \quad (10)$$

From Equation (10), the problem of deriving the conversion relationship between the camera pixel coordinate system and the projector pixel coordinate system can be converted into deriving the internal parameter matrix, the scaling factor, and the rotation matrix and translation vector of the camera and projector coordinate system.

According to the stereo calibration method, the camera's internal parameters and the $[R|T]$ matrix between the two camera coordinate systems can be easily obtained. However, because the projector cannot take pictures, deriving the projector's internal and external parameters requires the inverse simulation of the image at the projector's viewing angle. The specific processes are as follows:

1. Firstly, a black and white grid image P_p is projected on a blank plate, and camera A photographs the projected scene to obtain an image P_c . According to the reversibility of light, we can think of the image P_{p1} as the image taken by the projector on the projected scene;

2. Solving the homography matrix H_{cp1} between the images P_{p1} and P_{c1} based on the SURF matching algorithm [8]. Then the camera pixel coordinates and projector pixel coordinates have the following relationship:

$$\tilde{P}_p = H_{cp} * \tilde{P}_c \quad (11)$$

3. Keeping the position of the whiteboard unchanged, black and white checkerboards are placed on the whiteboard. The camera takes pictures of the checkerboard to obtain the image P_{c1} . In combination with the above-mentioned homography matrix H_{cp} , a corresponding “shooting” image of the projector is obtained. P_{p1} ;

4. After changing the position of the whiteboard, repeating the above steps (1) ~ (3) for seven times, to ensure the accuracy. As a result, we can not only obtain the grid image captured by the camera, but also obtained a group of projectors "shooting" image. Solves the problem that the projector cannot shoot;

5. According to these seven groups of images, the intrinsic parameters of the camera and the projector can be calculated as well as the rotation matrix R_{cp} and translation matrix T_{cp} between the two coordinate systems using the stereo calibration method.

It should be noted that, for a planar scene, the homography matrix is the transformation matrix between the camera and projector image pixel coordinate system. The solution of the relation matrix needs the projector to project the image each time. However, the assembly environment in real world is complex, and the projected surface is generally not a blank plate. Therefore, it is impossible to ensure the correct matching of images, not to mention the correctness of the homography matrix. Therefore, the method of directly deriving the homography matrix is not applicable to the actual situation in assembly. In this paper, step 3 to 5 in the procedure are proposed to solve the problem above. We only need to find the depth information s , then the dynamic characteristics of the system are guaranteed. The assembly operation can be continued while the assembly scene is changing.

2.2 Camera pose calibration based on natural feature points

Camera pose calibration is to get the rotation matrix and translation vector between the spatial coordinate system and the camera coordinate system. In order to get the transformation matrix between the camera and projector in real time, obtaining the position and orientation information of the camera in the space coordinate system in real time is needed.

PnP algorithm is an important algorithm to obtain the real-time pose information. However, the 2D coordinates and corresponding 3D coordinates of a series of points in the image must be given to run the algorithm. Typically, target scene will be labeled by 3-4 markers, the camera pose can be measured using those markers as well as both two-dimensional coordinate and three-dimensional coordinate. But placing markers will undoubtedly cause a lot of inconvenience. In some cases, it is impossible to

place markers. In addition, there are many variabilities during the assembly, making it difficult to ensure the same relative positional relationship between the camera-projector system and the assembly.

We proposed a solution that two cameras were used to photograph the assembly scene separately and automatically to obtain a series of natural feature points for the two images in the same scene. The relative positions of the projector and the two cameras are fixed. The physical system model is shown as follows in Figure. 5.

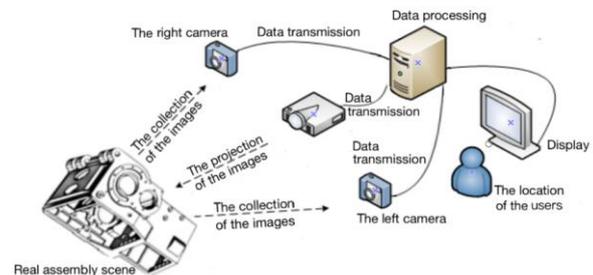


Figure 5. Physical model diagram of the combination system of dual cameras and single projector

The 2D coordinates of the natural feature points in the image can be easily obtained by using the Brief operator [9]. However it is hard to get the corresponding 3D coordinates for the reason that the depth information of the point can not be acquired in traditional way. Thereby the triangulation model is introduced here. The arranged cameras must meet the following three conditions:

1. The imaging planes are on the same plane, and the optical axes are strictly parallel and the focal lengths are the same.
2. The intersection c_l and c_r , which are the intersection of the two cameras' optical axis and the imaging plane have the same pixel coordinates in the image.
3. The pixel row of one camera is exactly the same as the other one, which means they have the same y-coordinate.

Assuming that the mapped points of point P in left and right image plane are p_l and p_r , their Abscissa are x_l and x_r . We define $d=x_l - x_r$ as disparity. Also, we name the depth of this point P to the camera's optical centre as Z and the distance between the left and right optical centers is T , as shown in Figure. 6:

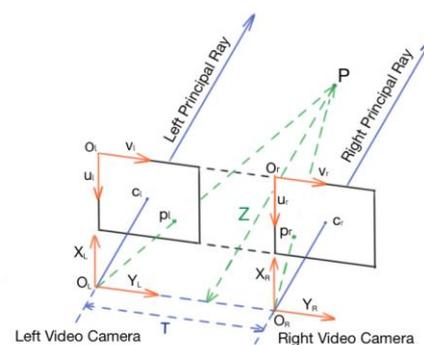


Figure 6. Triangulation principle

According to the principle of similar triangles, we can get the formula below:

$$\frac{T - (x_l - x_r)}{Z - f} = \frac{T}{Z} \quad (12)$$

The depth Z can be expressed as:

$$Z = \frac{fT}{x_l - x_r} \quad (13)$$

From the equation (9), we can know that in ideal situation, the depth of the spatial point is only related to the parallax, the focal length, and the distance between the optical centers of the two cameras. However, in actual situations, it is difficult to guarantee the two cameras to be strictly parallel. Therefore, the three-dimensional correction requires to be carried out before obtaining the depth information of the feature points, to get two aligned images with no distortion. Then the ORB feature matching is used to obtain the two-dimensional coordinates of the matching points in the two images. Combined with the intrinsic and external parameter matrices of the two cameras obtained in the previous section, the depth information of the corresponding points can be obtained.

After obtaining the two-dimensional coordinates of the natural feature points and the corresponding three-dimensional coordinates, the PnP algorithm can be used to complete the camera pose calibration. Combining formula (2) and the obtained related parameters, the camera scale factor can be obtained by multiplying and dividing the matrix, and all the parameters required by formula (10) are obtained.

Finally, according to the pose of the camera, the virtual and real image can be obtained. After the matrix transformation of formula (10) and remapping, the final virtual and real mixed image can be obtained.

3 Result

In this paper, we used the cable laying as an example to demonstrate this project. Two identical cameras and a miniature projector were used to form a camera-projector system. The physical system is as shown in the Figure. 7.

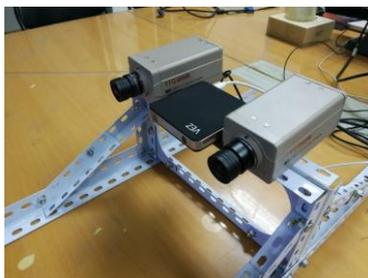


Figure 7. Projector-camera composite system physical map

Cable was laid on a flat panel in an experimental environment. The camera took an image PC. The solution proposed in this paper was used to reverse-solve the image PP “shot” by the projector and project the image into the cable laying scene. The image coincided with real scene, as shown in Figure. 8.

The left figure shows the calculated virtual model image projection on the real components. To show it

more clearly, the right figure shows the calculated virtual model image projection alone and the middle figure shows the corresponding real components. We can see them coinciding with each other very well.



Figure 8. Projection and real objects coincided with each other

To verify that the system could dynamically track changed in the projected scene, we adjusted the position of the projection plate in the experiment. The angle of the plane with the main projection surface was changed, and the rotations of 5° , 10° , and 15° were respectively tested. The above flow was repeated, and the image and the scene still overlapped, as shown in Figure. 9.



Figure 9. Image and real object after position changed

It can be seen that this experiment verified the effectiveness of the proposed scheme and realized the conversion of the camera field of view to the field of view of the projector. When the relative position between the real scene and the camera projector combination system changed, the system can complete the dynamic view. The field transformation proves the effectiveness of the system and verifies the effect of the dynamic adjustment of the system.

4 Conclusion

We proposed a method for converting the field of view between a camera and a projector based on binocular vision. Using this method, the projector's internal parameter matrix and the transformation relation matrix between the projector and the camera coordinate system were acquired, and the camera posed information was obtained by using natural feature points. Using this method made the camera-projector system more dynamic and practical. This article finally verified the effectiveness of the AR-assisted assembly system.

In the process of verifying the experimental results, when the rotation angle of the projection plane was too large, the projected image quality tended to be poor, and in the face of the occlusion problem, the implementation of the scheme was difficult. In the future, efforts will be made to solve the above two problems, and the system will be continuously improved. Finally, it will be applied to aerospace components assembly scenes.

Acknowledgment

This study was supported by Project of Equipment Pre-research (No. 41423010401).

References

1. Reiners D, Stricker D, Klinker G. Augmented reality for construction tasks: doorlock assembly[C]. International Workshop on Augmented Reality : Placing Artificial Objects in Real Scenes: Placing Artificial Objects in Real Scenes. A. K. Peters, Ltd. (1999):31-46.
2. YIN Xuyue, FAN Xiumin, WANG Lei, QIU Kai, LIU Rui. Augmented Reality Training System for Aerospace Product Assembly Process Guidance and Its Application[J]. *Aeronautical Manufacturing Technology*, (2018),61(Z1):48-53.
3. VVT. (2006). Augmented Assembly. Available: <http://www.vvt.fi/proj/augass/?lang=en>.
4. Yao Yuan. Studies on Implement Technology of Augmented Reality[D], Zhejiang University ,2006.
5. Chen Xiang-hui. Research on the Information Projection technique for Manual Assembly of Large Parts[D], Huazhong University of Science & Technology, (2016).
6. HA Yong-gang, ZHOU Ya, Wang Yong-tian, Yan Da-yuan. Head mounted display for augmented reality visualization[J]. *Optical Technique*,(2000), **04** :350-353.
7. Song Wei, Zheng Yi, Xie Xiang, Li Guo-lin. *An Accurate and Flexible Calibration Method for Projector-Camera*[J]. *Microelectronics & Computer*, (2015),32(2):29-32.
8. Herbert Bay , Andreas Ess , Tinne Tuytelaars , Luc Van Gool“SURF:Speeded up Robust Features”, *Computer Vision and Image Understanding*(CVIU),Vol.110,No.3,pp.346-359,(2006) .
9. M.Calonder,V.Lepetit,C.Strecha,and P.Fua,“Brief: Binary robust independent elementary features, ”in European conference on computer vision, pp.778-792, Springer,(2010).