Vison-based Auxiliary Navigation Method using Augmented Reality for Unmanned Aerial Vehicles

Huiyao Wu[1]
School of automation science and electrical engineering
Beihang University
Beijing, China
why.wuhuiyao@gmail.com

Zhihao Cai[2]
Institute of unmanned aerial vehicles
Beihang University
Beijing, China
caizhihao@gmail.com

Yingxun Wang[3]
School of automation science and electrical engineering
Beihang University
Beijing, China
wangyx@buaa.edu.cn

***Abstract*—A visual-based auxiliary navigation method using augmented reality is an immersive navigation interface with virtual navigation and decision-making information superimposed on the video of real environment. It can provide the operator with more information to behave proactively in order to forestall future problems, with less stress or effort, and to increase the successful rate of UAV missions. The components of the auxiliary navigation method using AR on UAVs are described in this paper. And a method of real-time displaying targets' position with virtual information based on computer vision registration is researched. The real-time information is obtained by the transformation between three coordinate systems. An experimental verification bases on ARToolkit platform that is able to identify multiple markers and display real-time virtual information of targets' distance and direction while generating auxiliary curves is also described.**

***Keywords*—augmented reality, navigation, UAVs, computer vision-based registration, ARToolkit**

I. INTRODUCTION

Augmented reality (AR) is a newly developed computer application and interaction technology. Unlike virtual reality (VR), where the user is completely immersed in a virtual environment, an augmented reality system is augmenting the real world scene. The computer-generated 3D graphics or 2D text is merged with the real view to help users to learn and perceive more information which cannot be seen in the real world [1]. It merges the latest achievements of digital image processing, computer graphics, artificial intelligence and multi-media technology, and is widely used for projects realization, entertainments, science researches and military affairs [2].

The field of military technology has paid considerable interest in augmented reality. Military personnel engaged in both combat and non-combat operations must understand a complex, dynamic environment, of which they often see only a small portion. AR, which is committed to combine computer-generated information with real-world scenes together, can provide the operator an accurate and efficient auxiliary interface, as well as an immersed interactive environment for training [3, 4].

Since presently, intelligent control and UAV platforms are difficult to achieve fully autonomous flight control, the operator of the UAV task execution results still have a great impact, especially when the UAV is in an unknown surroundings or performing tasks in harsh environments. Moreover, a survey on the general reliability of UAVs affirms that mishaps are due to the human on the ground in 17% of cases and suggests that this percentage could increase when considering the accidents and incidents where other system failures hide the influence of human–system design deficiencies [5]. So with virtual navigation and decision-making information superimposed on the video of real environment, the immersive navigation interface can provide the operator with more information to behave proactively in order to forestall future problems, with less stress or effort, and to increase the successful rate of UAV missions.

II. AUXILIARY NAVIGATION METHOD USING AR

A. The components and key technologies of AR system

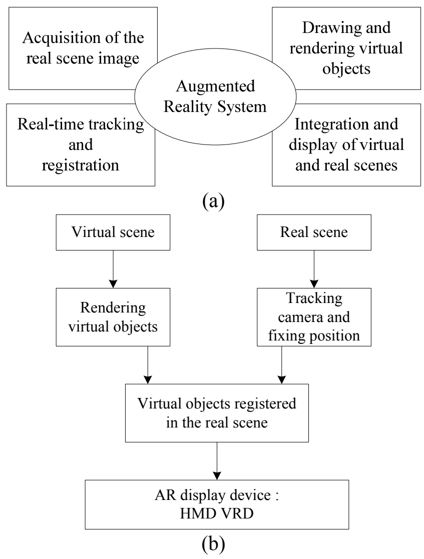

Figure 1. The components of AR system

978-1-4673-0311-8/12/$31.00 ©2012 IEEE

As showed in figure 1, the process of AR system is that getting the real world scene through camera, using real-time tracking technology to capture and calculate the camera's position and posture, then calculating the location of virtual objects according to the registration algorithm and the video superposed with virtual objects finally be displayed by the display technology.

Display technology and 3D registration technology are the key technologies of AR system. Display technology including: helmet display, ordinary display, hand-held display, projection display and so on. The registration is a process which merges virtual objects generated by computer with real world image caught by camera. It's is not only a key technology but also an important index for ensuring the performance of AR system [6].

In general, the registration technology can be classified into three kinds: tracker-based registration technology, knowledge-based registration technology and computer vision-based registration technology. Computer vision-based registration which is convenient to realization because of its easy theory is becoming a registration technology with high potential in the application of AR system. In this paper, registration based on the identification is used because of its simple equipment and high precision.

B. The Auxiliary Navigation Method using AR on UAVs

In the past, UAVs mainly rely on the Inertial Navigation System and Global Positioning System. However, INS faces the problem of accumulative error while GPS is not always available and the accuracy can't meet the need of UAVs sometimes [7].

The combination of augmented reality and visual navigation, can improve the navigation accuracy by the fusion of the real-time information provided by vision, INS and GPS, which makes up the defects [8].In the meantime, augmented reality can also be used to enhance the battlefield environment information. The integration of virtual objects in real environment, that is, superimposing the auxiliary information of navigation, environment and target, can create a highly realistic flight auxiliary navigation interface.

As shown in figure 2, the design strategy of auxiliary navigation method using AR on UAV includes the following five steps：

a) Get the real scene information.

b) Get the position and posture of UAV given by the navigation system.

c) Using image processing technology, find out the objects in the real scene video that matching the objects in feature library to be the target object.

d) Analyze the position and posture of the UAV and target object, and then use the vision-based registration algorithm to calculate the position that the virtual objects should superimposed on in real time.

e) Display the real scene and the virtual navigation information through the optical display helmet or video display [9].

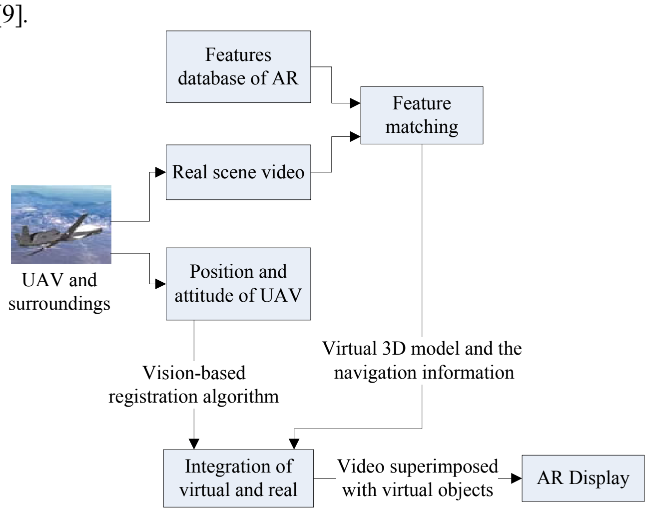

Figure 2. The Auxiliary Navigation Method using AR on UAVs

This paper will focus on vision-based registration method used on the auxiliary navigation system; design a simple augmented reality method using markers as the objects in feature library. Then the method of get the target objects' distance and direction information in real time using vision-based registration algorithm will be described in the following.

III. THE METHOD OF GET REAL-TIME INFORMATION OF DISTANCE AND DIRECTION

A. Computer vision-based registration

In order to merges virtual 3D objects seamless to the correct position of the real world in real time, vision-based tracking technology needs to obtain the position and direction of the camera relative to the real scene. Specifically, the registration is achieved by the transformation between three coordinate systems.

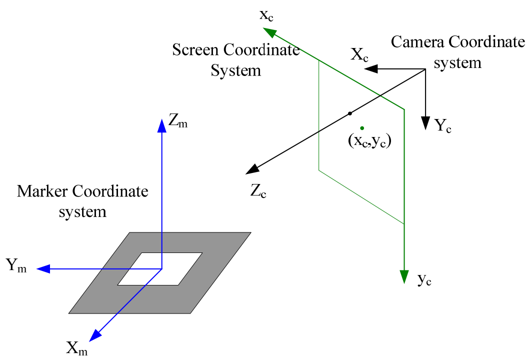

Figure 3. Marker coordinate system, camera coordinate system and screen coordinate system

On the computer screen, the row and column of each pixel is (u,v).Establish the screen coordinate in millimeters, a point in screen coordinate is (x_c, y_c) in figure 3. The pixel value of the base point in screen coordinate is (u_0, v_0) . Set the scale factor of x axis and y axis s_x and s_y $(pixel / mm)$, then the relationship between a pixel of an image and a point in screen coordinate system is [10]:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & u_0 \\ 0 & s_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \quad (1)$$

Then explain the relationship between the camera coordinate system and screen coordinate system. A point in camera coordinate system is (X_C, Y_C, Z_C) in figure 3. Using linear camera model (namely pinhole model), as shown in figure 4, the projection q of a point Q in the image is the intersection between the image plane and the line made of Q and the optical center.

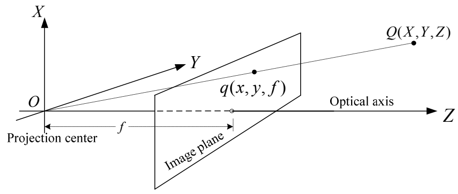

Figure 4. Linear camera models

Where point q in screen coordinate system is (x_c, y_c), and point Q in camera coordinate system is (X_C, Y_C, Z_C), the following matrix can be obtained by proportion:

$$\begin{bmatrix} Z_C x_c \\ Z_C y_c \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \quad (2)$$

Combine (1) and (2), we can get the transformation between pixels and camera coordinate system:

$$\begin{bmatrix} Z_C u \\ Z_C v \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} s_x f & 0 & u_0 & 0 \\ 0 & s_y f & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \quad (3)$$

Taking the distortion of the axis u, v into account, the actual equation is:

$$\begin{bmatrix} Z_C u \\ Z_C v \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} s_x f & k & u_0 & 0 \\ 0 & s_y f & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \quad (4)$$

Parameters s_x, s_y, f, u_0, v_0, k are only related to the internal structure of the camera, so they are called camera internal parameters, while matrix P is camera perspective matrix, which means the transformation matrix between the camera coordinate and the imaging plane. The perspective matrix P can be obtained by camera calibration. The camera's perspective matrix P in this paper is:

$$P = \begin{bmatrix} 484.44642 & 0.03134 & 363.50000 & 0.00000 \\ 0.00000 & 480.87293 & 188.00000 & 0.00000 \\ 0.00000 & 0.00000 & 1.00000 & 0.00000 \end{bmatrix}$$

Then discuss the process of the transformation between the three coordinate systems, that is, the method of registration. As the camera can be placed anywhere in the environment, a reference coordinate system is needed to describe the location of the camera, and any object in the environment, called the world coordinate system. The marker coordinate system is a world coordinate system whose base point locates on the marker. It consists of X_m, Y_m, Z_m axis. The relationship between camera coordinate system and marker coordinate system can be described by a rotation matrix R and a translation vector T. So, assume a point in space whose homogeneous coordinates in marker coordinate system and camera coordinate systems are $(X_m, Y_m, Z_m, 1)^T$ and $(X_C, Y_C, Z_C, 1)^T$, we get the following equation:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_W \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = M \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \quad (5)$$

M matrix is completely identified by the relative position and posture between the camera and the marker, so it is called the external parameter matrix of the camera, also called the projection matrix. Combine (4) and (5) we get:

$$\begin{bmatrix} Z_C u \\ Z_C v \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} s_x f & k & u_0 & 0 \\ 0 & s_y f & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = PMX \quad (6)$$

As expounded before, the perspective matrix P, which is fixed to a specific camera, can be obtained by camera calibration. The square of which the size is known and the internal pattern is asymmetry is used as the marker. So the coordinates of the four vertexes in marker coordinate system are known. Meanwhile, the coordinates of the four vertexes in screen coordinate system can be obtained in the image processing. In determining the perspective matrix P, based on four vertexes' coordinates in marker coordinate system and image coordinate system, the transition matrix between the two coordinate systems ,matrix M ,can be calculated, that is, the camera pose[11, 12].

B. The transformation based on registration algorithm to get distance and direction information

According to the known size of the markers and the registration algorithm, accurate distance non-contact measurement can be achieved, and the information can be superimposed on the target object in real time. The method will be discussed in the following.

First get the information of distance and direction between camera and the marker. The coordinate of a point in camera coordinate system is (X_C, Y_C, Z_C) ,expressed in C .In marker coordinate system the coordinate is (X_m, Y_m, Z_m) ,expressed in X .According to (5)，we can get:

$$\begin{bmatrix} C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix} = M \begin{bmatrix} X \\ 1 \end{bmatrix} \tag{7}$$

Now we need to know the marker's coordinate in camera coordinate system, that is to say, the point is in the center of the marker. Then $X = (0,0,0)$, and (7) turned into：

$$\begin{bmatrix} C \\ 1 \end{bmatrix} = M \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{8}$$

That is:

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} T_x \\ T_y \\ T_z \\ 1 \end{bmatrix} \tag{9}$$

It can be seen that marker's coordinate in camera coordinate system is the last column of the projection matrix M . And that's the reason vector T is called the translation vector. As the same argument the camera's coordinate in marker coordinate system can be obtained. When the point is on the camera lens, then $C = (0,0,0)$, we get:

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = M \begin{bmatrix} X \\ 1 \end{bmatrix} \tag{10}$$

$$\begin{bmatrix} C \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_1 \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} X_2 \\ 1 \end{bmatrix} \tag{11}$$

Thus it can be seen that camera's coordinate in marker coordinate system is the last column of the inverse matrix of the projection matrix M .Since the coordinates are obtained, the distance and direction information of markers and the camera can be got through a simple calculation.

Next discuss the method of get the distance and direction information between two markers. Assume that the two markers are marker 1 and marker 2. A point's coordinate in marker 1 coordinate system is $X_1 = (X_{m1}, Y_{m1}, Z_{m1})$;and its coordinate in marker 1 coordinate system is $X_2 = (X_{m2}, Y_{m2}, Z_{m2})$;its coordinate in camera coordinate system is $C = (X_C, Y_C, Z_C)$; the projection matrixes are M_1 and M_2，we get：

$$\begin{bmatrix} C \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_1 \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} X_2 \\ 1 \end{bmatrix} \tag{12}$$

Now we need to know the coordinate of marker 2 in marker 1 coordinate system, that is to say, the point is in the middle of marker 2，then $X_2 = (0,0,0)$:

$$M_1 \begin{bmatrix} X_1 \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{13}$$

$$\begin{bmatrix} X_1 \\ 1 \end{bmatrix} = M_1^{-1} M_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{14}$$

The coordinate of marker 2 in marker 1 coordinate system is the last column of the inverse matrix of marker 1's projection matrix M_1 multiply by marker 2's projection matrix M_2 . Since the coordinate of marker 2 in marker 1's coordinate system is obtained, the distance and direction information of the two markers can be got through a simple calculation.

IV. ARToolkit-based Experimental Verification

ARToolkit is a C language based on augmented reality system with secondary development package. It takes advantage of computer vision technology to calculate relative to the observer position and orientation of a known identity. It supports the integration of Direct3D, OpenGL graphics and VRML scenes to the video stream and also supports the enhancement based on visual or video reality applications [13].

The following this paper bases on ARToolkit platform which uses markers as target objects, designs an auxiliary navigation method that is able to identify multiple markers and display real-time virtual information of targets' distance and direction while generating auxiliary curves.

The markers that represent the target objects in this paper are as following:

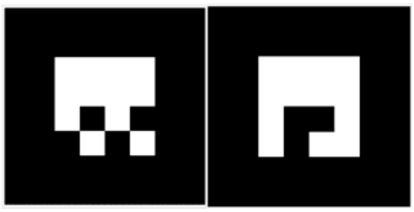

Figure 5. Markers in this paper

Binarization processing according to the threshold shows as below:

Figure 6. Binarization processing of the markers

The 3D virtual models in this paper are as following:

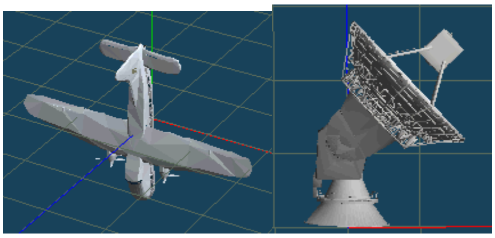

Figure 7. The 3D virtual models

The final effect of the method is shown in figure 8, 9, 10. When the target objects (i.e. markers) that match the objects in augmented reality library appear in the video the camera captures, according to the registration technology in 3.1, the virtual objects and information will be superimposed on the corresponding position in the real video scene. And distance and direction information can be displayed according to the transform method in 3.2. As shown in Figure 9 and 10, the distance of virtual information decreases (distance in mm, the error within 1mm), orientation is also changing as the camera get closer to the targets. Auxiliary decision is displayed when the distance is within a certain range.

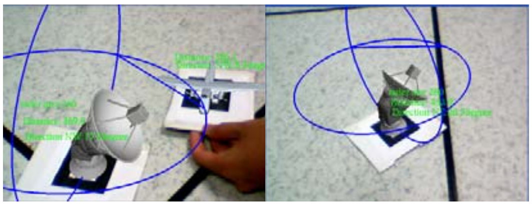

Figure 8. The overall effect

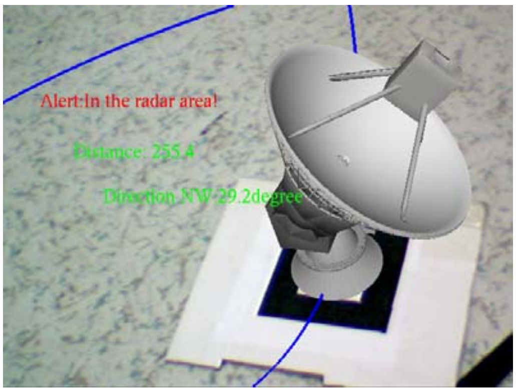

Figure 9. The camera get close to the objects to be avoided

(a) The distance is 704.8mm,and the direction is northwest 5.6degree

(b) The distance is 382.2mm,and the direction is northwest 41.8degree

(c) The distance is 207.9mm,and the direction is northwest 40.7 degree .

Display the decision-making information of "in attack range"

Figure 10. The process of the camera get close to the target object

V. CONCLUSION

This paper puts forward combining augmented reality technology and vision-based navigation to design an auxiliary navigation method with strong sense of reality. It can increase the rate of UAV successful missions with virtual navigation and decision-making information superimposed on the video of real environment. This paper analyzes the vision-based registration technology and put forward the method of getting distance and direction information in real time, finally experimental verifies on ARToolkit platform.

From the experimental results it can be seen that the method has achieved good results in the case of that specific markers represent the target object indoor. But further outdoor applications require the identification of more complex objects, so only the binary image processing is not enough. At the same time the identification and superposition of more objects requires more works on real-time (no delay), stability (accurate, no vibration) and robustness (is not affected by light, shelter and motion). These issues, pending further study.

REFERENCES

[1] R. Azuma, "Recent advances in augmented reality PB -," IEEE Computer Graphics and Applications, vol. 21, pp. 34-47, 2001.

[2] R. T. Azuma, "A Survey of Augmented Reality," Teleoperators and Virtual Environments, vol. 6, pp. 355-385, 1997.

[3] L. Kisung, K. Yongkwon, I. C. Seong, and C. Kyungho, "Building Detection in Augmented Reality Based Navigation System," *1611-3349,* 2007-07-07 2007.

[4] M. A. Livingston, "User interface design for military AR applications PB -," Virtual Reality*,* vol. 15, pp. 175-184, 2011.

[5] F. De Crescenzio, G. M. F. Persiani and T. Bombardi, "A First Implementation of an Advanced 3D Interface to Control and Supervise UAV (Uninhabited Aerial Vehicles) Missions," Virtual Reality and Simulation, vol. 3, 2009.

[6] Y. B. Li, "Development Actuality and Application of Registration Technology in Augmented Reality PB -," Proceedings of The 2008 International Symposium on Computational Intelligence and Design, vol 2, pp. 69-74, 2008.

[7] G. Raphael, M. Alessandro, B. Mark, and S. Dieter, "Navigation Techniques in Augmented and Mixed Reality: Crossing the Virtuality Continuum," 978-1-4614-0064-6*,* 2011-08-30 2011.

[8] W. Cheng,X J Liao,Z. Li, " Study on Augmented Reality and Its Military Application" Computer Technology and Development, pp. 165-168, 2007.

[9] Z. Cai, "Multi-source information fusion augmented reality benefited decision-making for unmanned aerial vehicles: A effective way for accurate operation PB -," Proceedings of the 2011 6th IEEE Conference on Industrial Electronics and Applications, ICIEA 2011, vol. ISSU, pp. 174-178, 2011.

[10] K.Shen, " Research of the Human-Computer Interaction System for Augmented Reality Based on ARToolkit ". vol. Master: Hefei University of Technology, 2007.

[11] M. B. S. W. Hirokazu Kato, "A Mixed Reality 3D Conferencing Application,", 2000.

[12] K.Huang, " Research and Application on the Algorithms of Camera Calibration on Computer Stereo Vision,". vol. Master: Shenyang University of Technology, 2007.

[13] J.Yi ,Z.L.Li, " Stereovision Imaging Augmented Reality System Based on ARToolKit" Microcomputer Information, pp. 255-257, 2007.