IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 28, 2018, accepted August 5, 2018, date of publication August 7, 2018, date of current version August 28, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2864224

A Real-Time Three-Dimensional Tracking and Registration Method in the AR-HUD System

ZHE AN[1], XIPING XU[1], JINHUA YANG[1], YANG LIU[1], AND YUXUAN YAN[2]
[1]School of Optoelectronic Engineering, Changchun University of Science and Technology, Changchun 130022, China
[2]State Key Laboratory of High Power Semiconductor Lasers, Changchun University of Science and Technology, Changchun 130022, China

Corresponding author: Xiping Xu (xxp@cust.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Project 61605016 and in part by the National Key Scientific and Technological of China under Project 20160204062GX.

ABSTRACT In this paper, we propose a method of 3-D tracking registration with a virtual image and the surroundings in three steps: initial registration matrix acquisition, feature extraction in a road scene, and registration of virtual information in a real environment. We obtain the initial registration matrix according to the relationship between different parts of the system. During transformation matrix solving, we do not match features. Instead, the relationship is based on the characteristic image information. Then, we interpolate another image's corresponding coordinates and estimate the transform matrix of the map and the image. The average time of the virtual reality registration algorithm can reach 0.0781 s. The average accuracy of our registration method is less than 2 pixels. The proposed algorithm has some advantages. We use features from accelerated segment test feature extraction and the binary robust independent elementary features descriptor to efficiently extract image feature points. The algorithm is more efficient, because it eliminates the feature-matching phase. We apply the proposed virtual reality registration method to a vehicle's augmented reality head-up display (AR-HUD). We derive the relationship between human eyes, optical system, tracking camera, and real scene in the AR-HUD system. Our experiment proves that the algorithm has a better registration effect on virtual objects, and it enhances road information. Different from the general head-up display (HUD) system, our method achieves the true integration of virtual and real world. It improves the distraction problem caused by the mismatch between the existing HUD virtual image and the real scene.

INDEX TERMS Augmented reality, virtual-real registration, camera pose estimation, AR-HUD.

I. INTRODUCTION

With the progress of science and technology, augmented reality (AR) has gradually come into people's field of vision. AR adds virtual information to real surroundings to give people more useful information [1]–[6]. Pokémon games and heads-up display (HUD) systems on vehicles are typical examples of AR technology. The key problem with AR is the registration of virtual information, that is, how to stack the virtual information generated by a computer into a real environment according to the surroundings [7]–[10]. AR consists of several parts: an optical display system, a tracking and registration mechanism, and real-time interaction. Holographic technology can solve the integral imaging based on visual fatigue problems to capture a number of two-dimensional (2D) images from two different angles of a three-dimensional (3D) scene by using a micro lens array and charge coupled device (CCD) sensor or camera array [11]. Hong *et al.* [10] and Li *et al.* [13], [14] have implemented an augmented reality system using full color holographic lenses and lens array functions, that can provide full color image in a 3D virtual optical transmission AR system. In this study, the AR-HUD system, which is different from existing visual display methods and is based on the free-form surface of a reflector, is taken as the object of study. We applied tracking registration with virtual image and real scene to a virtual road sign projected onto the AR-HUD. This is aimed at effectively improves traffic safety.

In order to achieve AR, some HUD systems have been proposed. Tonnis [15] studied the problems in the development of AR systems, and analyzed the common development process of the auxiliary system in the driving simulator. This was aimed at effectively improving traffic safety, and a new solution of the AR technology to combine virtual images with surroundings was proposed based on a context analysis and inference subsystem for spatial sensor data. Ng-Thow-Hing *et al.* [16] focused on drivers' three important items: understanding driver perception, driver distraction, and driver behavior. They created and evaluated driving solutions using augmented reality.

In the AR-HUD system, tracking registration of the virtual image and real scene is an important issue. Several methods have been proposed [17]–[21]. In terms of target-tracking, the traditional method is unstable and is greatly affected by noise, especially if the object can be easily lost during movement.

Skrypnyk and Lowe [22] proposed a method based on scale-invariant feature transforms (SIFT) features. The method uses features extracted by SIFT to match features between adjacent frames. Then, to complete the registration, it uses feature point information to obtain the transformation matrix. In the detection of target feature points, the SIFT algorithm can obtain reliable feature points with scale invariance. Unfortunately, this method requires extensive processing time, such that it has an impact on the 3D registration system, making it difficult to achieve real-time performance.

In practical applications, multiple tracking and registration techniques are combined [23]. Researchers at the University of Southern California [24] used a hybrid tracking method that combines inertial sensors with model-based visual tracking registration for an outdoor environment. The Spain Paloc team [25] combined flat visual inspection and an inertial sensor with tracking and location. However, the sensor was easily inhibited by the external environment.

To address unknown scene tracking registration, Davison *et al.* [26] proposed a registration method for a large-scale simultaneous localization and mapping (SLAM) method and developed the world's first online tracking system based on monocular vision. It was applied to an AR registration problem with unknown scenes. This method uses filtering technology to solve the problem of structured scenes, but it can only be used to reconstruct and track a few natural feature points online.

Subsequently, other scholars proposed series of multiple schemes, such as parallel tracking and mapping (PTAM) [27], large scale direct (LSD)-SLAM [28], and ORB-SLAM [29]. Among these schemes, the ORB-SLAM turned out to be a very stable system. It is constructed from sparse feature points, which can meet location demands for scene tracking. Based on this foundation, Salas-Moreno *et al.* [30] proposed a dense SLAM method using semantic information. This method performs semantic recognition and classification of a scene. The semantic plane structure obtained by the algorithm can be effectively utilized. They applied this approach to AR systems, including browsing virtual image content on walls or other planes. This method increases pixel-level information at the object level, which is a relatively new method and produces a better registration effect.

Our method is different from previous methods owing to the following sections. Our contribution is that we proposed a method to throw the virtual image accurately to the driver's line in the road scene.In this paper, we propose a new method of virtual registration. The registration accuracy is less than 2 pixels. The method is robust and eliminates the time and steps of feature matching, and thus obtains better registration results than previous methods. Finally, the true fusion of the virtual image and the real scene is realized. We apply the algorithm with AR-HUD to improve the safety of driving process, and the driving safety is improved.

II. PROPOSED METHOD

A. SYSTEM ARCHITECTURE

The system architecture of the proposed method is shown in Fig. 1.

We mainly investigated the technology involving tracking registration, which, in the case of AR-HUD movement, determines how to accurately stack virtual images generated by a computer into a real scene and then displays them to the driver's eyes. Therefore, in the first step, we obtained the initial registration matrix acquisition by established a relationship between the camera, driver's eyes, real objects' positions, and the virtual plane where the real objects are projected.

In the second step, we extracted the features from a real scene. To obtain accurate registration, the exact relationship between the virtual information obtained from the current scene and the camera must be guaranteed. That is, we must ensure accurate tracking. To realize the correct registration of virtual information, the key problem to solve is the estimation of the transformation matrix of the camera at different moments. By this method, we can determine the extent of the camera's change in viewpoint and calculate the registration location [31], [32]. The image taken by the camera is regarded as a matrix. In the process of pose acquisition, we do not make use of the entire picture; instead, we extract feature points from the scene to reduce the amount of calculation. We extracts the features of a road scene using features from accelerated segment test (FAST) and binary robust independent elementary features (BRIEF) methods. After extracting key points from the two images, we use the fast library for approximate nearest neighbors (FLANN) algorithm [33] to match the image feature points. Then use the random sample consensus (RANSAC) [34] method to eliminate redundant points during matching. So the transformation matrix is obtained according to the characteristic relationship between camera images.

The third step consists of estimating the camera pose. We build an initial map library that is used for estimating the camera pose with image frames. Then, we match the binocular images taken by the left and right cameras, and we transform the image points into spatial points. Then, we find the relation model between the spatial points and the feature points from the map library. Unlike in previous methods that match two sets of feature points first and then use the iterative closest point (ICP) or perspective from n points (PnP) methods to solve the camera pose, we omitted the feature-matching procedure. Instead, we modelled their relationship. Finally, the camera pose matrix was obtained by solving the model. We used OpenGL to obtain the image results. In this way, the driver can see the virtual image after registration.

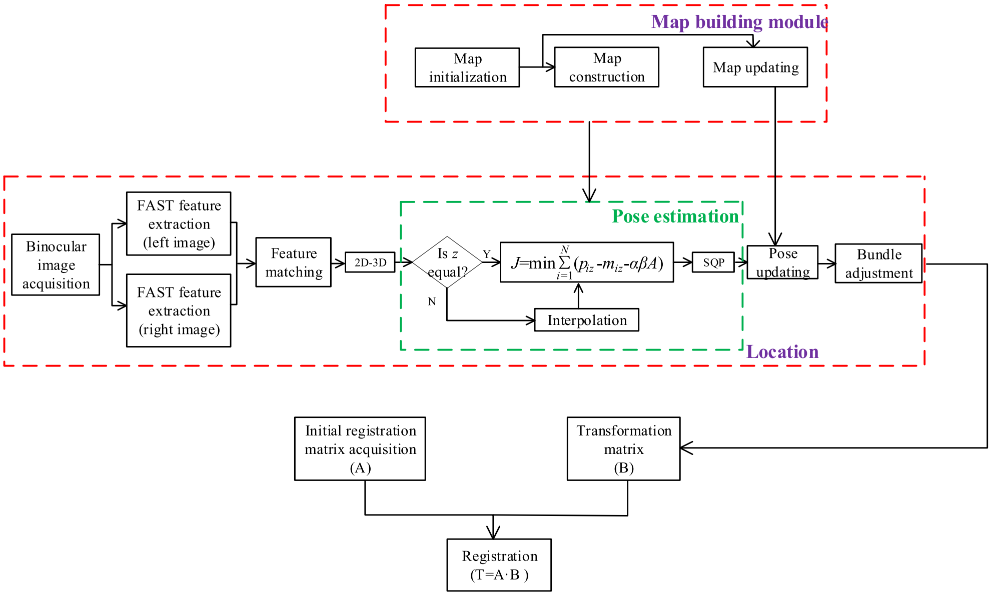

Map building module
Map initialization
Map construction
Map updating
Binocular image acquisition
FAST feature extraction (left image)
FAST feature extraction (right image)
Feature matching
2D-3D
Pose estimation
Is z equal?
Y
N
J=min $\sum_{i=1}^{N}(p_{iz}-m_{iz}-\alpha\beta A)$
SQP
Interpolation
Pose updating
Bundle adjustment
Location
Initial registration matrix acquisition (A)
Transformation matrix (B)
Registration (T=A·B)

FIGURE 1. Workflow of the proposed approach.

B. INITIAL REGISTRATION MATRIX ACQUISITON

AR-HUD is a kind of auxiliary driving system based on AR technology, which uses optical projection to present virtual information enhancements in the driver's line-of-sight without frequent adjustment between the real world and dashboard/navigation data. The working principle is shown in the following figure.

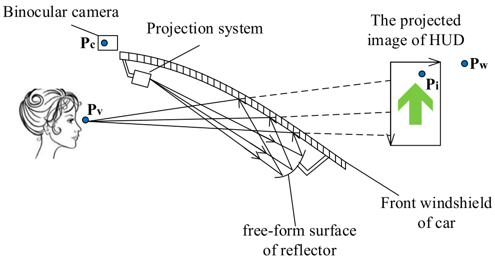


FIGURE 2. Working principle of AR-HUD.

where $\mathbf{p_w}$ is the any point in space, $\mathbf{p_c}$ is the corresponding 3D point of $\mathbf{p_w}$ on the binocular camera, $\mathbf{p_v}$ is the mapping points in human eye coordinate system, $\mathbf{p_i}$ is the point of $\mathbf{p_v}$ in projected image.The relationship between them can be represented by the following formula. The human eye observes the virtual image through the imaging system, and both constitute a pinhole model virtual camera, as shown in Figure 2. The virtual image can be defined through the virtual coordinate system.

$$\mathbf{P_i} = \mathbf{K_v^{3\times4}}\begin{pmatrix}\mathbf{r}^{3\times3}_{v\leftarrow c} & \mathbf{t}^{3\times1}_{v\leftarrow c}\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{r}^{3\times3}_{c\leftarrow w} & \mathbf{t}^{3\times1}_{c\leftarrow w}\\ 0 & 1\end{pmatrix}\mathbf{P_w} \tag{1}$$

The variables involved are: **r** is the rotation matrix between coordinate systems, **t** is the translation matrix between coordinate systems, and $\mathbf{K_v}$ represents the internal parameters of a virtual camera. In this way, the initial registration matrix **A**(**r**,**t**) can be obtained by solving **r**,**t**.

C. FEATURE EXTRACTION

To obtain a suitable road image, a binocular camera is used to obtain the scene image. The road scenario considered in this study is an unknown scenario of real-time change. We used the FAST method to extract the key points, which mainly detects obvious local pixel gray change. FAST is a type of corner point, which mainly detects the obvious change of the local pixel gray level. The speed of this method is fast. The idea is that if a pixel is different from the neighbourhood pixel, then it may be a corner point. The speed was faster than other detection algorithms, and the detection was too quick to compare brightness differences between pixels. To make the detected corner points have a description of scale and rotation, the gray centroid method can be used to describe these two properties. After extracting the key points, we used BRIEF descriptor to calculate each point such that the extracted features maintain good effect in translation, rotation, and scaling transformation.

The binocular camera takes two images at each moment. We use the FAST method to extract the feature points of these images, then we match their feature points and eliminate the redundant points. The matching feature leverages the coordinate points of the image plane between the left and right cameras, $\mathrm{C_1}$ and $\mathrm{C_2}$. We define the two points as $\mathrm{p_1}$ and $\mathrm{p_2}$. In their respective images, the corresponding coordinates are $\mathbf{p_1} = (u_1, v_1, 1)^T$, $\mathbf{p_2} = (u_2, v_2, 1)^T$. Suppose $\mathbf{p_1}$ and $\mathbf{p_2}$ correspond to the same spatial point coordinate, $\mathbf{P}(x, y, z)$. The homogeneous coordinate of point $\mathbf{P}$ in the world coordinate system is $\mathbf{P} = (x, y, z, 1)^T$. The projection matrices in the calibrated left and right camera were

N_1 and N_2, respectively. We obtained the following relationships.

$$Z_{C1}\begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = \begin{pmatrix} n^1_{11} & n^1_{12} & n^1_{13} & n^1_{14} \\ n^1_{21} & n^1_{22} & n^1_{23} & n^1_{24} \\ n^1_{31} & n^1_{32} & n^1_{33} & n^1_{34} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2}$$

$$Z_{C2}\begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} = \begin{pmatrix} n^2_{11} & n^2_{12} & n^2_{13} & n^2_{14} \\ n^2_{21} & n^2_{22} & n^2_{23} & n^2_{24} \\ n^2_{31} & n^2_{32} & n^2_{33} & n^2_{34} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{3}$$

By eliminating the parameter Z_{C1} and Z_{C2} in formular (2) and (3), the following equations are obtained.

$$\begin{cases} (u_1 n^1_{31} - n^1_{11})x + (u_1 n^1_{32} - n^1_{12})y \\ \quad + (u_1 n^1_{33} - n^1_{13})z = n^1_{14} - u_1 n^1_{34} \\ (v_1 n^1_{31} - n^1_{21})x + (v_1 n^1_{32} - n^1_{22})y \\ \quad + (v_1 n^1_{33} - n^1_{23})z = n^1_{24} - v_1 n^1_{34} \end{cases} \tag{4}$$

$$\begin{cases} (u_2 n^2_{31} - n^2_{11})x + (u_2 n^2_{32} - n^2_{12})y \\ \quad + (u_2 n^2_{33} - n^2_{13})z = n^2_{14} - u_2 n^2_{34} \\ (v_2 n^2_{31} - n^2_{21})x + (v_2 n^2_{32} - n^2_{22})y \\ \quad + (v_2 n^2_{33} - n^2_{23})z = n^2_{24} - v_2 n^2_{34} \end{cases} \tag{5}$$

Because the spatial point **P** satisfies both formular (4) and formular (5), we combined the equations to obtain a single set of four equations with three unknown variables, which should have unique solutions, and the "least squares" method was used to solve them. Then, the 3D coordinates of point **P** in space were obtained. We can estimate the camera's pose by processing the above-mentioned feature points at each moment.

D. REGISTRATION OF VIRTUAL IMAGE AND REALITY SCENE

After obtaining the 3D coordinates of the image frame, the transform matrix between the map and the image frame can be estimated. The usual method is to match the feature points first, and then use ICP or PnP to compute the transformation matrix. To reduce the algorithm computation time and simplify the steps, we did not match the feature points; instead, we estimated the camera pose directly by establishing a modeled relationship between feature points. First, we established the relationship between the map's and the image's feature points, which is the coordinate of two groups of points in space at the same point on a surface as the corresponding point (if there are no corresponding points on the interpolation of a temporary point). The spatial coordinates corresponding to the image feature points were set as $\mathbf{p_i}(\mathbf{p_x}, \mathbf{p_y}, \mathbf{p_z})$; then, we applied the difference between square and minimum principle to the z coordinates of corresponding points to establish the objective function. After rotation and translation, the corresponding coordinate of $\mathbf{p_i}$ is $\mathbf{p'_i}$, obtained by sequential quadratic programming (SQP) used to solve the conversion parameter vector and estimate the transformation matrix. The relationship is defined as

$$\mathbf{p'_i} = \mathbf{Rp_i} + \mathbf{T} \tag{6}$$

Here, **R** is the rotation matrix between $\mathbf{p_i}$ and $\mathbf{p'}_i$, and **T** is thetranslation matrix between $\mathbf{p_i}$ and $\mathbf{p'}_i$. Then, the "least squares" problem was constructed to make the sum of squares error reach the minimum **R** and **T**, and to establish the relationship between $\mathbf{p_i}$ and $\mathbf{m_i}$.

$$J = \min\sum_{i=1}^{N}\left\|\mathbf{p'_i} - \mathbf{m_i}\right\|^2 = \min\sum_{i=1}^{N}\left\|\mathbf{p'}_i - (\mathbf{Rp'_i} + \mathbf{t})\right\|^2 \tag{7}$$

We used the first-order Taylor method to expand formular (7) so that

$$\mathbf{p'_i} = \mathbf{R_0 p} + \mathbf{t_0} + d\mathbf{R}\cdot\mathbf{p_i} + d\mathbf{t} \tag{8}$$

where $d\mathbf{R} = \begin{pmatrix} 0 & -d\gamma & d\beta \\ d\gamma & 0 & -d\alpha \\ -d\beta & d\alpha & 0 \end{pmatrix}$ is a small change in the rotation matrix, $d\mathbf{t} = (dt_x, dt_y, dt_z)^T$ is a small change in the translation matrix.

Because $\mathbf{p'_{i0}} = \mathbf{R_0 p} + \mathbf{t_0}$, the formular (8) can be expressed as

$$\mathbf{p'}_i - \mathbf{p'_{i0}} = d\mathbf{R}\cdot\mathbf{p_i} + d\mathbf{t} \tag{9}$$

If $\mathbf{p_i}$ passes through, the **R**,**T** transform produces a small change of $d\mathbf{p_i}$ and their relationship can be expressed as $d\mathbf{p_i} = \mathbf{p'_i} - \mathbf{p'_{i0}}$. Substituting in formular (9), the following relationship was obtained.

$$d\mathbf{p_i} = \begin{pmatrix} 0 & -d\lambda & d\phi \\ d\lambda & 0 & -d\omega \\ -d\phi & d\omega & 0 \end{pmatrix}\cdot\begin{pmatrix} p_{ix} \\ p_{iy} \\ p_{iz} \end{pmatrix} + \begin{pmatrix} dt_x \\ dt_y \\ dt_z \end{pmatrix} \tag{10}$$

The above equation can be expanded to a component-wise form:

$$\begin{cases} dp_{ix} = -p_{iy}d\lambda + p_{iz}d\phi + dt_x \\ dp_{iy} = p_{ix}d\lambda + p_{iz}d\phi + dt_y \\ dp_{iz} = -p_{ix}d\phi + p_{iy}d\omega + dt_z \end{cases} \tag{11}$$

Taking the x and y coordinates of the image space points as the basis, the search map points at the same x,y coordinates under z, are $p'_{ix} = m_{ix}$, $p'_{iy} = m_{iy}$. If the z coordinates are equal, we record this point. If not, the z coordinates corresponding to x and y are interpolated in the map library and defined as m_{iz0}. This is used instead of the original coordinates for the camera pose estimation. At this point, formular (7), defines a point that lies between the image's feature points and the point of the map becomes

$$\begin{aligned} J &= \min\sum_{i=1}^{N}\left\|\mathbf{p'_i} - \mathbf{m}_i\right\|^2 \\ &= \min\sum_{i=1}^{N}\left[(p'_{ix} - m_{ix})^2 + (p'_{iy} - m_{iy})^2 + (p'_{iz} - m_{iz})^2\right] \\ &= \min\sum_{i=1}^{N}(p'_{iz} - m_{iz})^2 \end{aligned} \tag{12}$$

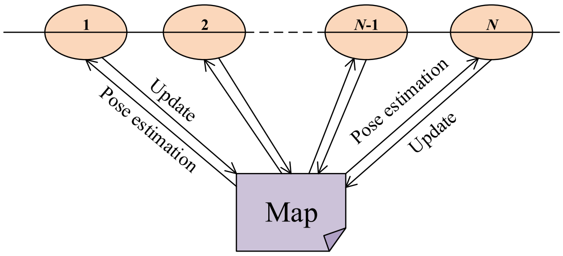


FIGURE 3. Operational principle of camera pose estimation.

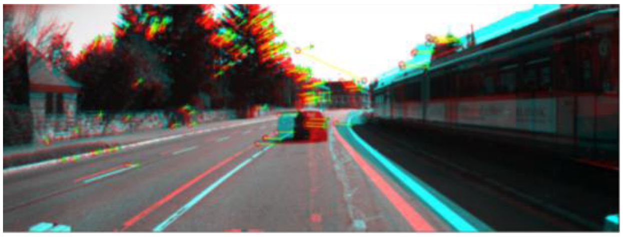


FIGURE 4. Result of feature-matching.

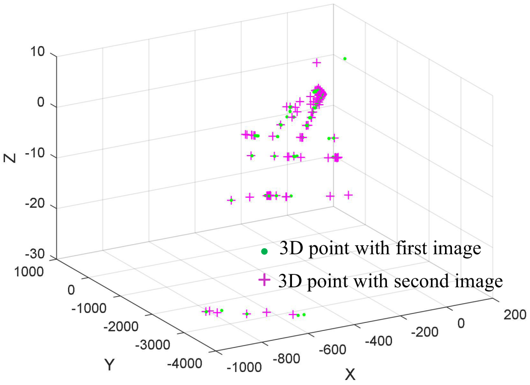


FIGURE 5. Conversion result of 2D-3D points.

provided the feature point coordinates of the map library satisfy the following:

$$z_i = f(x_i, y_i) = z_0 + dz_i = z_0 + \frac{\partial f_i(x,y)}{\partial x_i} dx_i + \frac{\partial f_i(x,y)}{\partial y_i} dy_i \qquad (13)$$

Here, the components $d\mathrm{p}_{\mathrm{ix}}$ and $d\mathrm{p}_{\mathrm{iy}}$ of the $d\mathbf{p_i}$ should be equal to the $d\mathrm{x_i}$ and $d\mathrm{y_i}$ of formular (12), so

$$m_{iz} = m_{iz0} + \frac{\partial f_i(x,y)}{\partial x_i} dp_{ix} + \frac{\partial f_i(x,y)}{\partial y_i} dp_{iy} \qquad (14)$$

Substituting formular (14) and the ordinate component p'_{iz} of $\mathbf{p_i}$ into formular (12), we obtain the following equation:

$$J = \min \sum_{i=1}^{N} (p_{iz} - m_{iz} - \alpha \boldsymbol{\beta} \boldsymbol{B})^2 \qquad (15)$$

where $\alpha = (\frac{\partial f_i(x,y)}{\partial x_i}, \frac{\partial f_i(x,y)}{\partial y_i}, 1)$ is the normal vector of $\mathrm{m_i}$,

$$\boldsymbol{\beta} = \begin{pmatrix} 0 & -p_{iz} & p_{iy} & 1 & 0 & 0 \\ p_{iz} & 0 & p_{ix} & 0 & 1 & 0 \\ p_{iy} & -p_{ix} & 0 & 0 & 0 & 1 \end{pmatrix}$$

is a construction matrix, and $\mathbf{B} = (d\omega, d\phi, d\lambda, dt_x, dt_y, dt_z)^T$ is the parameter vector. So we've got the camera's transformation matrix. The final registration matrix is a synthesis of initial matrix $\mathbf{A}$ and transformation matrix $\mathbf{B}$:

$$\mathbf{W} = \mathbf{A} \cdot \mathbf{B} \qquad (16)$$

The feature of the first frame image was extracted and all the extracted feature points were inserted into the map library. The map refers to the set of feature points that were cached from the feature points of each frame. In the process of using the map, each frame either adds some feature information or it updates the location estimation of older feature points. We modeled the current image and map points, and obtained the estimation results of the camera pose. Each image frame contributes some characteristic information to the map. we can obtain a precise map that ensures accurate camera pose estimation. The operational principle is shown in Fig. 3.

In the experiment, the sensor must have a certain noise, it may also be affected by the magnetic field and temperature, which leads to the error of estimation. The error of each estimation will be accumulated to the next estimation, which leads to inaccurate estimation of the long time matrix. Therefore, the Bundle Adjustment (BA) method is used to deal with the accumulated errors. Next, we integrated the virtual images with the real environment. The virtual camera's internal parameters and external parameters were defined before this process was carried out. Then, the texture was used to draw the real image onto the plane perpendicular to the virtual camera's viewpoint. Then, the virtual object was projected between the plane and the camera. Owing to the projection process, we obtained virtual fusion scenes in the imaging plane.

III. EXPERIMENTAL RESULTS

The proposed approach was implemented using Linux on a 2.9 GHz CPU with 8 GB RAM. We used the KITTI public data set for algorithm experiments. First, the initial map library was built. Then, the feature points were extracted from the binocular images and matched. We use the RANSAC method to eliminate redundant points during matching. The results are shown in Fig. 4.

Using this approach, we obtained the matched image feature points. Based on the method described in Sec. C, the feature points were transformed into spatial points and compared with the map feature points, as shown in Fig. 5. The feature points (red crosses) in the map are very close to the image feature points (green points), so that most points appear violet in color. Based on the image feature points, we searched for each corresponding point on the map to find its x and y coordinates. If the corresponding z value was different,

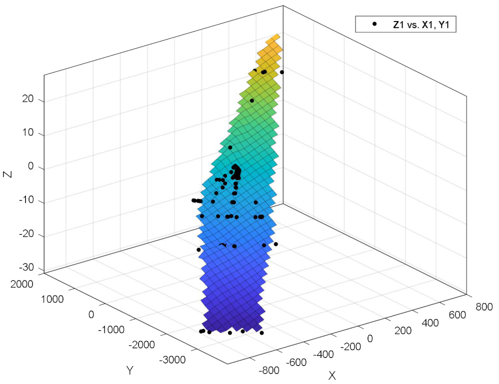


FIGURE 6. Interpolated images of map feature points.

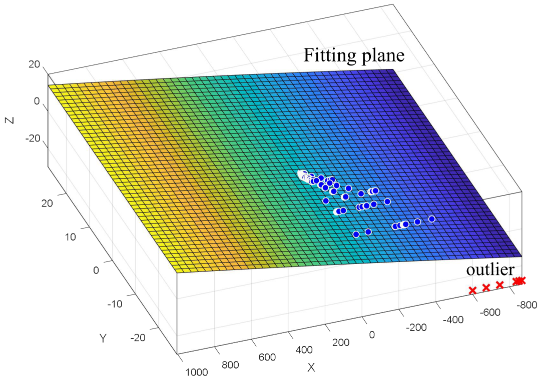


FIGURE 7. Fitting images of map feature points.

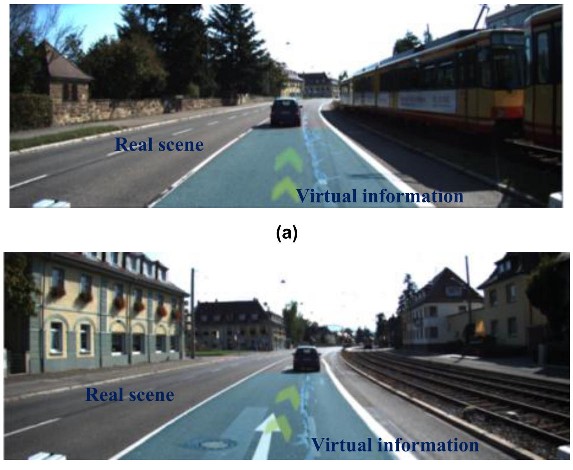
TABLE 1. Results of linear constraint of SQP.

Iteration number	Number of function calls	Iteration step	Iterative error
0	3	-	2.87e+04
1	13	10	1.03e+03
2	21	7.6646	115
3	27	2.338534	0.05243
4	31	0.0003759	4.88e-05

the m_{iz0} point was interpolated. The interpolation map of the feature points is shown in Fig. 6.

In terms of α in formular (15), a surface was fitted to the map point. The fitting surface could be a plane or another surface, depending on the distribution of data. By requesting the partial values of the partial differential equations at $\mathrm{x_i}$ and $\mathrm{y_i}$, the fitting results are as shown in Fig. 7.

The blue origin is the characteristic point, and the red crosses are points where the fitting error is relatively large and should be eliminated. After calculating the parameter values in fomular (15), the SQP method was used to solve the nonlinear constraint problem. The results of the linear constraint are presented in Table 1.



(a)

(b)

FIGURE 8. Images of virtual-real registration of different times.

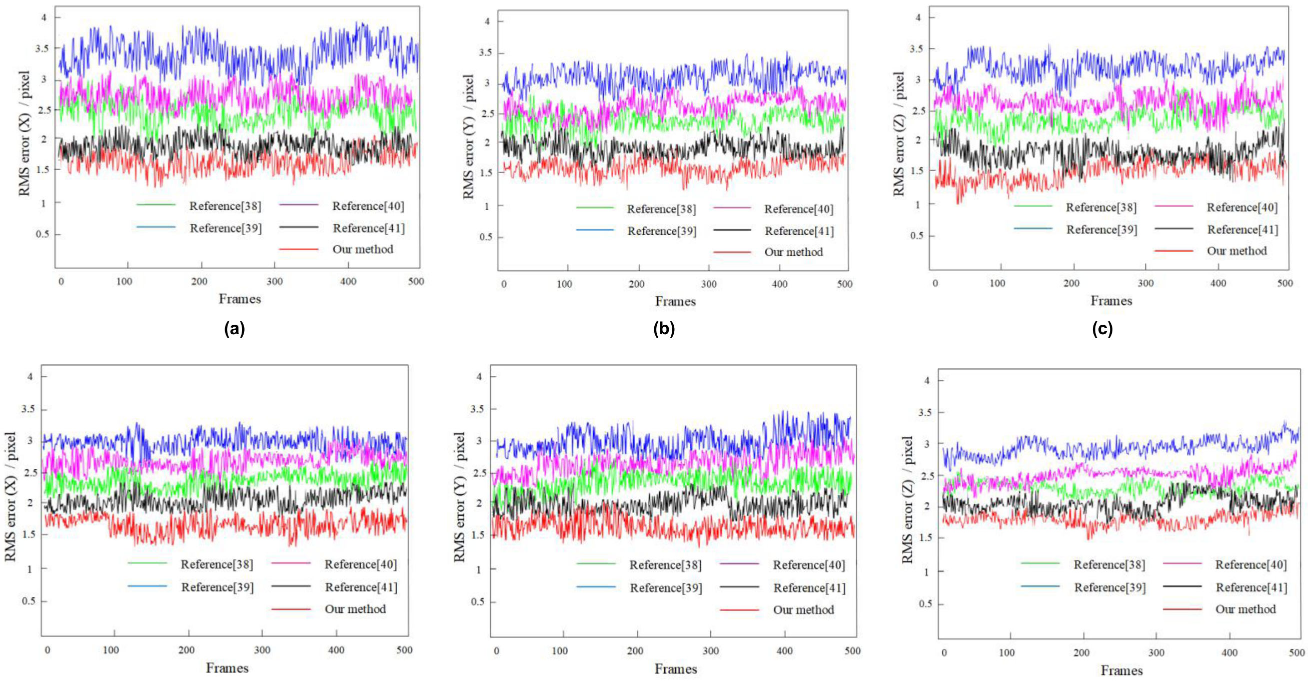
Based on the data from the above table, we can estimate the camera's transformation matrix after four iterations. We compared our algorithm with the actual registration method based on PnP [35], ICP [36] and the triangulation [37]. Table 2 presents the average time spent in each step and the total time spent by each methods. The feature points were estimated by simplifying the algorithm steps, which shortened the time of the entire process. Thus, the algorithm meets the real-time requirements of a virtual-real registration system.Once the transformation matrix of the camera in tense is obtained, the virtual image can be superimposed onto the real scene by formular (16). Experiments were conducted to register the scene from different perspectives, and the results are shown in Fig. 8 below.

Although our method and other algorithms are based on the least squares principle, there is no time consuming preprocessing step. But there's calculation method of camera pose estimation is different. The point distance in the algorithm is the Z coordinate difference between different points. The algorithm established in this paper is relatively simple, and the corresponding relationship is established. The computation is small and the computation efficiency is very high. Other methods have higher requirements for establishing point correspondences, and the speed of convergence is fast. The number of iterations is very close to the real value. Howerver, the large number of computation points and the overall computation efficiency is low. Generally speaking, the pose difference between matching points is not well. Therefore, the proposed algorithm is suitable for three-dimensional tracking and registration registration.

From the principle of the algorithm, the FAST feature extraction method can quickly extract the features of different road scenes. Moreover, in the camera pose estimation processing, we combine the feature extraction results of the previous frame images. In this way, we have more features. We can prove the accuracy of the algorithm in different views. Moreover, the algorithm doesn't need to match the images,

TABLE 2. Comparison of average processing times.

Steps	Average processing time (seconds)					
	Based on PnP		Based on ICP		Triangulation	Proposed method
Initial Registration Matrix Acquisition(A)	0.0257					
Binocular feature processing	0.0278		0.0261		0.0239	0.0221
Pose estimation(B)	Feature-matching	0.0046	Feature-matching	0.0047	0.0315	0.0283
	PnP	0.028	ICP	0.031		
Registration ($T = A \cdot B$)	0.002					
Total time (seconds)	0.0881		0.0895		0.0831	0.0781



(a) (b) (c)

(d) (e) (f)

FIGURE 9. The figure of RMS error. (a) rotate around X-axis. (b) rotate around Y-axis. (c) rotate around Z-axis. (d) move along X-axis. (e) move along Y-axis. (f) move along Z-axis.

and also improves the speed of the algorithm. Experimental results demonstrate that the algorithm can adapt to different perspectives and different road conditions, and that the registration results are stable. In the experiment, the rotation error of the camera on X, Y and Z axis is recorded, and the movement along the X,Y and Z axis is carried out to test the tracking accuracy of the algorithm under rotation,translation and scale changes. By calculating the RMS error between the real coordinates of the feature points and the re-projection coordinates of the virtual-real registration matrices, the performance of the registration algorithm is tested in this study and compared with the two methods with better registration effect at present. We tested 500 images and Fig. 9 shows the error comparison of these methods.

Figure 9 shows that the registration error of the proposed method is smaller than [38]–[41]. The registration accuracy of the algorithm in four states is less than 2 pixels. The main source of false and real registration errors in AR-HUD system is the error caused by jitter during vehicle running. Therefore, we will reduce the jitter error in future.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a tracking registration method to place virtual objects into a realistic scene.

First, we obtained the initial registration according to the coordinate transformation matrix. Then, we extracted the image features, and the 2D points were transformed into spatial points according to the feature points of the binocular images. However, when the camera matrix was estimated, feature-matching was omitted and the transformation camera matrix was obtained directly. Thus, the registration of the virtual image in the real environment was realized. Experimental results showed that the algorithm speed is improved compare with the other registration methods. The improved method not only had a better registration effect, but was stable and satisfied the requirements for the registration in a real-world road scenario. Therefore, our approach can be applied to a vehicle's AR-HUD.

Our experimental results showed that our algorithm can accurately and rapidly complete the registration of a virtual image and real surroundings in real time and can adapt to different scene angles. However, this study has limitations. During the experiment, we found that if the camera produces a slight jitter, the depth of the feature point estimated was subject to error. To overcome this problem, we placed feature points on the map and optimized them as a group. Using this approach, the registration result obtained was more accurate and stable.

Another limitation is maintaining a continuously updated map. The arrival of new images per frame will increase the number of map features. We believe that removing features from the field will reduce the burden on the computer. However, it is easy to lose features because an environment affected by light is not ideal for extraction. In the case of severe occlusion, features can easily be lost. Inaccurate pose estimation can also directly lead to inaccurate registration. Therefore, solutions to these problems will be the focus of our future work.

REFERENCES

[1] K. Khairnar, K. Khairnar, S. Mane, and R. Chaudhari, "Furniture layout application based on marker detection and using augmented reality," *Int. Res. J. Eng. Technol.*, vol. 2, no. 7, pp. 540–544, 2015.

[2] C. Santos, T. Araújo, J. Morais, and B. Meiguins, "Hybrid approach using sensors, GPS and vision based tracking to improve the registration in mobile augmented reality applications," *Int. J. Multimedia Ubiquitous Eng.*, vol. 12, no. 4, pp. 117–130, 2017.

[3] U. Rehman and S. Cao, "Augmented-reality-based indoor navigation: A comparative analysis of handheld devices versus Google glass," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 1, pp. 140–151, Feb. 2017.

[4] J. Joo-Nagata, F. M. Abad, J. G.-B. Giner, and F. J. García-Peñalvo, "Augmented reality and pedestrian navigation through its implementation in m-learning and e-learning: Evaluation of an educational program in chile," *Comput. Educ.*, vol. 111, pp. 1–17, Aug. 2017.

[5] B. Kress and T. Starner, "A review of head-mounted displays (HMD) technologies and applications for consumer electronics," *Proc. SPIE*, vol. 8720, p. 87200A, May 2013.

[6] M. Maisto, C. Pacchierotti, F. Chinello, G. Salvietti, A. De Luca, and D. Prattichizzo, "Evaluation of wearable haptic systems for the fingers in augmented reality applications," *IEEE Trans. Haptics*, vol. 10, no. 4, pp. 511–522, Oct./Dec. 2017.

[7] A. Maimone, A. Georgiou, and J. S. Kollin, "Holographic near-eye displays for virtual and augmented reality," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–16, 2017.

[8] S. Fani, S. Ciotti, E. Battaglia, A. Moscatelli, and M. Bianchi, "W-FYD: A wearable fabric-based display for haptic multi-cue delivery and tactile augmented reality," *IEEE Trans. Haptics*, vol. 11, no. 2, pp. 304–316, Apr./Jun. 2018.

[9] J. Li, J. A. Besada, A. M. Bernardos, P. Tarrío, and J. R. Casar, "A novel system for object pose estimation using fused vision and inertial data," *Inf. Fusion*, vol. 33, pp. 15–28, Jan. 2017.

[10] Ó. Blanco-Novoa, T. M. Fernández-Caramés, P. Fraga-Lamas, and M. A. Vilar-Montesinos, "A practical evaluation of commercial industrial augmented reality systems in an industry 4.0 shipyard," *IEEE Access*, vol. 6, no. 6, pp. 8201–8218, 2018.

[11] X.-W. Li and I. K. Lee, "Modified computational integral imaging-based double image encryption using fractional Fourier transform," *Opt. Lasers Eng.*, vol. 66, pp. 112–121, Mar. 2015.

[12] K. Hong, J. Yeom, C. Jang, J. Hong, and B. Lee, "Full-color lens-array 494 holographic optical element for three-dimensional optical see-through augmented reality," *Opt. Lett.*, vol. 39, no. 1, pp. 12–130, 2014.

[13] X. Li, D. Xiao, and Q.-H. Wang, "Error-free holographic frames encryption with CA pixel-permutation encoding algorithm," *Opt. Lasers Eng.*, vol. 100, pp. 200–207, Jan. 2018.

[14] X.-W. Li, S.-T. Kim, and Q.-H. Wang, "Copyright protection for elemental image array by hypercomplex Fourier transform and an adaptive texturized holographic algorithm," *Opt. Express*, vol. 25, no. 15, pp. 17076–17098, 2017.

[15] M. Tönnis, "Towards automotive augmented reality," Univ. Munich, Munich, Germany, Tech. Rep., 2008.

[16] V. Ng-Thow-Hing, K. Bark, L. Beckwith, C. Tran, R. Bhandari, and S. Sridhar, "User-centered perspectives for automotive augmented reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality-Arts, Media, Humanities*, Oct. 2013, pp. 13–22.

[17] T. Hayashi, H. Uchiyama, J. Pilet, and H. Saito, "An augmented reality setup with an omnidirectional camera based on multiple object detection," in *Proc. 20th Int. Conf. Pattern Recognit.*, Istanbul, Turkey, Aug. 2010, pp. 3171–3174.

[18] S.-H. Kong *et al.*, "Robust augmented reality registration method for localization of solid organs' tumors using CT-derived virtual biomechanical model and fluorescent fiducials," *Surgical Endoscopy*, vol. 31, no. 7, pp. 2863–2871, 2017.

[19] Y. J. Wang, P.-J. Chen, X. Liang, and Y.-H. Lin, "Augmented reality with image registration, vision correction and sunlight readability via liquid crystal devices," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 433.

[20] J. Park, B.-K. Seo, and J.-I. Park, "Binocular mobile augmented reality based on stereo camera tracking," *J. Real Time Image Process.*, vol. 13, no. 3, pp. 571–580, 2017.

[21] B. Streckel and R. Koch, "Lens model selection for visual tracking," in *Proc. Joint Pattern Recognit. Symp.* Berlin, Germany: Springer, 2005, pp. 41–48.

[22] I. Skrypnyk and D. G. Lowe, "Scene modelling, recognition and tracking with invariant image features," in *Proc. 3rd IEEE ACM Int. Symp. Mixed Augmented Reality*, Nov. 2004, pp. 110–119.

[23] T. Oskiper, S. Samarasekera, and R. Kumar, "Multi-sensor navigation algorithm using monocular camera, IMU and GPS for large scale augmented reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, Atlanta, GA, USA, Nov. 2012, pp. 71–80.

[24] B. Jiang, U. Neumann, and S. You, "A robust hybrid tracking system for outdoor augmented reality," in *Proc. IEEE Virtual Reality*, Chicago, IL, USA, Mar. 2004, pp. 3–275.

[25] I. Barandiaran and C. Paloc, "Hybird tracking for outdoor augmented reality applications," *Comput. Graph. Topics*, vol. 18, no. 4, pp. 6–15, 2006.

[26] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.

[27] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. 6th Int. Symp. Mixed Augmented Reality*, Nara, Japan, Nov. 2007, pp. 225–234.

[28] E. Jakob, S. Thomas, and C. Daniel, "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. Eur. Conf. Comput. Vis.*, Zürich, Switzerland, 2014, pp. 834–849.

[29] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.

[30] R. F. Salas-Moreno, B. Glocken, P. H. J. Kelly, and A. J. Davison, ''Dense planar SLAM,'' in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2014, pp. 157–164.

[31] P. Furgale, J. Rehder, and R. Siegwart, ''Unified temporal and spatial calibration for multi-sensor systems,'' in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Tokyo, Japan, Nov. 2013, pp. 1280–1286.

[32] S. Weiss and R. Siegwart, ''Real-time metric state estimation for modular vision-inertial systems,'' in *Proc. IEEE Int. Conf. Robot. Automat.*, Shanghai, China, May 2011, pp. 4531–4537.

[33] L. L. Zhao, G.-H. Geng, K. Li, and A.-J. He, ''Images matching algorithm based on SURF and fast approximate nearest neighbor search,'' *Appl. Res. Comput.*, vol. 30, no. 3, pp. 921–923, 2013.

[34] R. Schnabel, R. Wahl, and R. Klein, ''Efficient RANSAC for point-cloud shape detection,'' *Comput. Graph. Forum*, vol. 26, no. 2, pp. 214–226, 2007.

[35] Y. Wu and Z. Hu, ''PnP problem revisited,'' *J. Math. Imag. Vis.*, vol. 24, no. 1, pp. 131–141, 2006.

[36] J. Chen, ''University N F. application of ICP algorithmin TLS deformation monitoring,'' *Mod. Surveying Mapping*, vol. 40, no. 5, pp. 5–7, 2017.

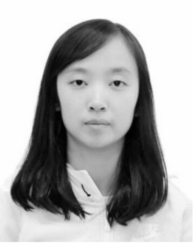
[37] E. Marchand, H. Uchiyama, and F. Spindler ''Pose estimation for augmented reality: A hands-on survey,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 12, pp. 2633–2651, Dec. 2016.

[38] L. Yu, S. K. Ong, and A. Y. C. Nee, ''A tracking solution for mobile augmented reality based on sensor-aided marker-less tracking and panoramic mapping,'' *Multimedia Tools Appl.*, vol. 75, no. 6, pp. 3199–3220, 2016.

[39] T. Fei, L. Xiao-Hui, H. Zhi-Ying, and H. Guo-Liang, ''A registration method based on nature feature with KLT tracking algorithm for wearable computers,'' in *Proc. Int. Conf. Cyberworlds.*, Sep. 2008, pp. 416–421.

[40] B. Xiao, Y. Jiang, and F. Lin, ''Mobile augmented reality 3D registration algorithm based on ORB and KLT,'' *Comput. Modernization*, vol. 3, no. 3, pp. 57–60, 2014.

[41] Z. Zhang and Y. Liu, ''Tracking and AR registration based on SURF and optical flow method,'' *Comput. Eng. Appl.*, vol. 3, no. 3, p. 035, 2015.



ZHE AN received the B.S. degree from the Changchun University of Science and Technology, Jilin, China, in 2016, where she is currently pursuing the Ph.D. degree with the School of Optoelectronic Engineering Technology. Her research focus is on image recognition and tracking.

XIPING XU received the bachelor's degree in electronic engineering from the Changchun Optics Precision Mechanical School in 1995 and the master's and Ph.D. degrees from the Changchun University of Science and Technology in 1999 and 2004, respectively. He is currently a Professor and a Doctoral Supervisor with the Changchun University of Science and Technology. His research interests include photoelectric detection technology and quality control.

JINHUA YANG received bachelor degree from the Xian University in 1990, and obtained master's degree from the Beijing Institute of Technology University in 1997 and doctoral degree in 2001 in Changchun University of science and technology. Now he is a professor and a doctoral supervisor in Changchun University of Science and Technology, and his research interests include photoelectric detection technology and quality control.

YANG LIU received the B.S. degree from the Changchun University of Science and Technology, Jilin, China, in 2016, where he is currently pursuing the Ph.D. degree from the School of Optoelectronic Engineering Technology. His research interests include optical communication and the field of coherent detection.

YUXUAN YAN received the bachelor's degree from the Changchun University of Science and Technology in 2016, where he is currently pursuing the master's degree in electronic science and technology. His main research interest is physics and technology of semiconductor lasers.