IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 11, 2019, accepted February 24, 2019, date of publication March 13, 2019, date of current version March 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902188

H∞ Optimal Control-Based Robust Pose Estimation in Light Field Three-Dimensional Display

ZHENHAO WANG, YAN ZHAO, AND SHIGANG WANG
College of Communication Engineering, Jilin University, Changchun 130012, China
Corresponding author: Yan Zhao (zhao_y@ jlu.edu.cn)

This work was supported in part by the National Key Research and Development Plan under Grant 2017YFB1002900 and Grant 2017YFB0404800, in part by the National Natural Science Foundation of China under Grant 61631009 and Grant 61771220, and in part by the Fundamental Research Funds for the Central Universities under Grant 2017TD-19.

ABSTRACT Naked eye augmented reality technology has the capability of satisfying the immersion feeling. Therefore, light field three-dimensional (3-D) display technology has received considerable attention. In this paper, a novel pose estimation method in light field 3D display based on H∞ optimal control is proposed. Compared with the current pure hardware platform method, the proposed method releases hardware resources such as the graphic processing unit and avoids numerous parameter adjustments. The proposed method comprises three key parts. First, we calculate the normal of the rigid objects and apply it to obtain the normal estimation. Second, we calculate H∞ of the normal estimation result on the Hardy space, which is the maximum singular value of the rational function matrix parsed in the right half plane of the complex plane. Finally, we use the optimal result of H∞ for pose estimation and transformation. The experiments are carried out on the virtual and real classical datasets acquired by Kinect, Mian, and Clutter. The experiments show that the proposed method can be used to obtain high-quality display effects with low cost and high efficiency.

INDEX TERMS H∞, pose estimation, light field 3D display.

I. INTRODUCTION

Naked eye three-dimensional (3D) augmented reality technology is widely used in various fields, such as demonstration-teaching, exhibition displays, media videos, and medical treatments. Different from traditional display technology, the naked-eye true 3D augmented reality technology has unique characteristics that do not require viewers to wear glasses or helmets to watch the 3D effects. Its realistic depth of field and 3D sense greatly enhances the visual impact and immersion of the audience during the viewing experience. Consequently, naked-eye 3D technology is the best choice of display product for promotion, publicity, and video playback. Nowadays, 3D display technologies are mainly classified as binocular, integrated imaging, volumetric, holographic, and light field.

Previously, the holographic 3D display method was thought to be the only one. However, with the improvement of high-definition pixels, high modulation rates of various light modulators, and rapid developments in computer image graphics technology; the 3D display of a light field of directional light reconstructed by a geometric optical orientation screen and projection technology, has gradually emerged. Compared with the traditional holographic display, the 3D display of the light field can produce a high-quality and high-resolution color dynamic 3D display [1]. It also displays complex textures and lighting shadows [2]–[5]. Therefore, the study of the 3D display of the light field has recently received considerable attention. Zhan *et al.* [6] proposed a novel light field display system by implementing liquid crystal Pancharatnam-Berry phase lenses to relieve the discomfort from the convergence-accommodation conflict.

Since the development of light field 3D display technology, the 3D display of the light field methods mainly implement real-time updating and dynamic processing of the 3D object from acquisition to display, according to the hardware adjustment. This includes the data compression algorithm and parallel operation of the graphic processing unit (GPU).

The associate editor coordinating the review of this manuscript and approving it for publication was Yichuan Jiang.

A multiple viewpoint rendering (MVR) algorithm is presented in [7] that produces the track and texture mapped scene renderings, well-suited for hardware acceleration. Hou *et al.* [8] made use of the NVIDIA Geforce-7800-GT graphics card chip, which for GPU rendering, represents each triangle on the simplified mesh as a simple multi-perspective camera. The aim of the method was to pre-process the correspondence between the original surfaces for a dense sample of 3D locations and to store the result within the joint texture map. Nevertheless if the culling is in the 3D viewing frustum, then the problem would be much more complicated for efficient GPU implementation. Reference [9] also used GPU vertex shaders for the multiple center of projection (MCOP). This is a hardware algorithm to view the light field display rendering, and the hardware DVI output of the graphics card was the NVIDIA GeForce 8800 and an FPGA-based image decoder. Notwithstanding, such hardware platforms are not practical, due to their age. Although the core graphics and image projection components decrease in price, the capabilities do not increase. The image processing chips with more advanced capabilities already exist, but their prices are more expensive.

In recent years, the accuracy and interactive ability of the 3D light field display has gradually improved [10]–[13]. Hand gesture recognition [14], identification of pedestrians [15], face recognition, and other artificial intelligence functions are quickly combined. Some new algorithms have also emerged in this field [16]–[20]. However, hardware equipment costs and complexity have also increased. Navarro and Buades [13] proposed a new matching method, which computes the disparity maps between specific pairs of views for estimating the depth from a light field image. Notwithstanding, inaccurate registrations led to an increase in the resolution of light field images. Ferreira and Gonçalves [18] introduced a novel algorithm similar to RANSAC that generates a coarse depth map with one depth from the 3D point cloud, to two depths, and a surface fitting per micro-lens with different focal-lengths. The paper [21] designed a point cloud coding solution for the geometry of static point clouds using octree- and graph-based transforms, which led to the application scenarios addressing lower to medium rates, and a coding solution with a better performance. A new light field depth estimation method is presented in [19], which locates the optimal orientation of an epipolar plane image and local linear embedding. Nevertheless, this approach is applicable only for a light field with a small number of views. Anisimov and Stricker [22] presented an algorithm for the depth estimation from light field images in a relatively small amount of time, using only a single thread on the CPU that combines stereo matching and line fitting approaches. To the best of our knowledge, this is the first time that a method has been developed that contains no 3D light field display algorithm implemented by GPU. With the increasing popularity of virtual reality and augmented reality technology, 3D light field display technology combined with computer vision and machine learning, can provide a viable path for producing low-cost, high-quality VR and AR content [23]. Incidentally, the paper [24] presented a comprehensive overview regarding the research of light field technology over the last 20 years.

In this paper, we present an innovative method for obtaining the vertex position coordinates by the H∞ optimal control of the robust pose estimation of rigid objects during 3D light field image processing. The advantages of the proposed method include convenience, efficiency, a low cost, and less error. The proposed method is based on mathematics and can be implemented only using a single-threaded CPU. This releases hardware resources such as GPU and avoids manual error adjustments of the parameters, ultimately improving the efficiency and accuracy.

The remainder of the paper is organized as follows. Section 2 describes the proposed method in detail. Section 3 presents the experimental results and Section 4 provides the conclusion.

A block diagram of the whole light field acquisition-display data processing based on the proposed method, is illustrated in Fig. 1. The proposed strategies are denoted in purple. First, there is downsampling, followed by the normal and feature estimations. The mathematical core of the H∞ optimal control is then performed, which includes the maximum singular value as obtained by the Jordan singular value decomposition (SVD). Then, the rotation matrix and translation vector are used to align the point cloud. Finally, we obtain the accuracy vertex position coordinates and robust pose estimation of rigid objects in the light field 3D imaging. Table 1 describes the detailed algorithm of the H∞ optimal control based robust pose estimation of rigid objects in the light field 3D display.

II. APPROACHES

A. NORMAL ESTIMATION

The normal vector reflects the first-order differential properties of the surface of the object, which can more accurately describe the surface features represented by the point cloud. Therefore, we first find the normal vector estimation based on this. In point-based graphics, the point-based precision and high-quality rendering mainly depend on the normal vector information, and many surface reconstruction algorithms also require the use of the normal vector to obtain reliable reconstruction results. For reconstruction algorithms that require normal vector aggregation, the detection and recovery of sharp features relies entirely on reliable and accurate normal vectors. If the correct normal vector can be calculated, even geometric features of point clouds with severe defects can be perceived.

The local neighborhood fitting method was firstly proposed by Hoppe *e*t al. [25]. In the tangent plane, the point p is obtained by the least square fitting of the k nearest neighbor of the point p, and the principal component analysis (PCA) is used to solve the covariance matrix of the neighborhood point. The eigenvector corresponding to the minimum eigenvalue of the covariance matrix is the normal vector of

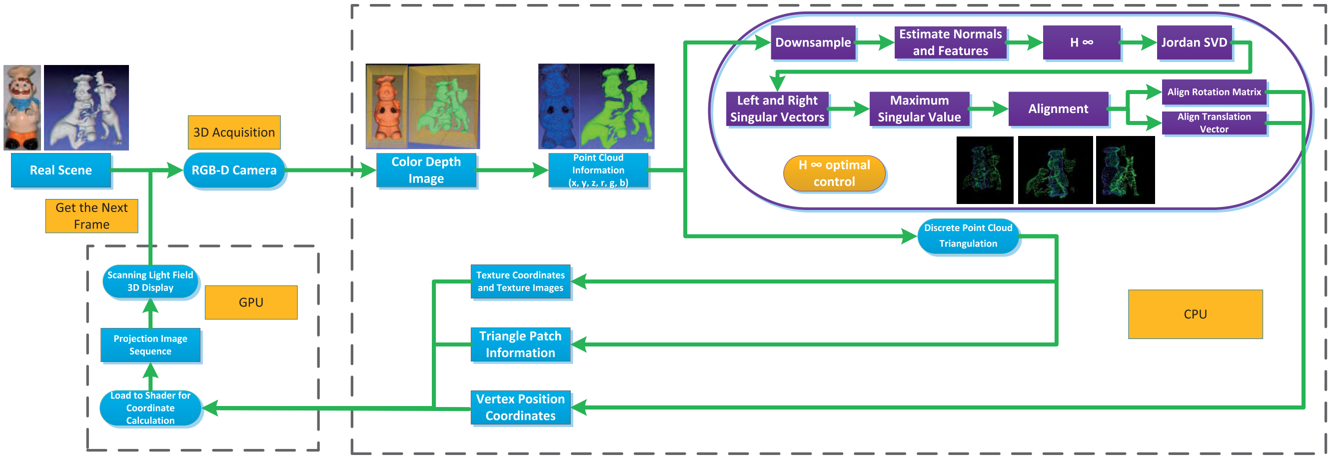


FIGURE 1. **Block diagram of the whole light field acquisition-display data processing based on the proposed method.**

TABLE 1. **Algorithm for the H∞ optimal control based robust pose estimation of rigid objects.**

ALGORITHM : H∞ optimal control based robust pose estimation of rigid objects in light field 3D display

Begin

1) Use RGB-D camera to acquire 3D real scene.

Input(CPU unit):

2) Load the color and depth image of the 3D real scene and collect the point cloud information.

3) Get the vertex position coordinates by H∞ optimal control.

4) Make use of the discrete point cloud triangulation to obtain the texture coordinates and triangle patch information.

Input(GPU unit):

5) Use the above three variables output from the CPU as the input variables of the GPU, which is loaded into the shader for the coordinate calculation.

6) Transform to the projection image sequence.

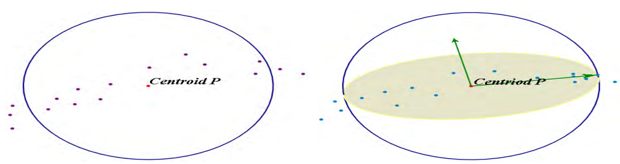
7) Scan light field 3D display.

Loop:

8) Get the next frame.

End

the plane, and is also the normal of the point p, as shown in Fig. 2. The PCA normal estimation method is used for the normal estimation in this paper.



(a) (b)

FIGURE 2. **Schematic diagram of the point cloud local covariance analysis. (a) K nearest neighbors selection. (b) Covariance analysis.**

The point cloud is a set of $P = \{p_1, p_2, \ldots, p_n\}$ and n is the total point cloud number. The nearest k neighbor of point p_i is denoted by $N_b(p_i)$, where N_b represents the neighborhood. The representation of the least square plane fitted to its k neighbors for any point p_i is given by

$$Pl(n, d) = \operatorname{argmin} \sum_{p_i \in N_b} (np_i - d)^2 \tag{1}$$

where n is the normal vector of the plane of Pl, n must satisfy $\|n\|_2 = 1$; d is the distance from the neighborhood point to the fit plane. The equation (1) is transformed into an eigenvalue decomposition for the semi-positive covariance matrix C in equation (2), in which the eigenvector of the minimum eigenvalue of C is used as the normal vector of p_i. Fig. 3 shows the virtual and realistic dataset application results for the normal estimation. As shown in Fig. 3, the effect and versatility of the normal vector estimation in this paper presented good results for various scenes.

$$C = \begin{bmatrix} p - p_1 \\ p - p_2 \\ \vdots \\ p - p_k \end{bmatrix}^T \begin{bmatrix} p - p_1 \\ p - p_2 \\ \vdots \\ p - p_k \end{bmatrix} \tag{2}$$

B. H∞

The H∞ is vital in this paper, which is the principle of pose estimation. In this paper the signal space for light field imaging is $L_p[0, \infty)$, $1 \leq p \leq \infty$, and its extended point cloud

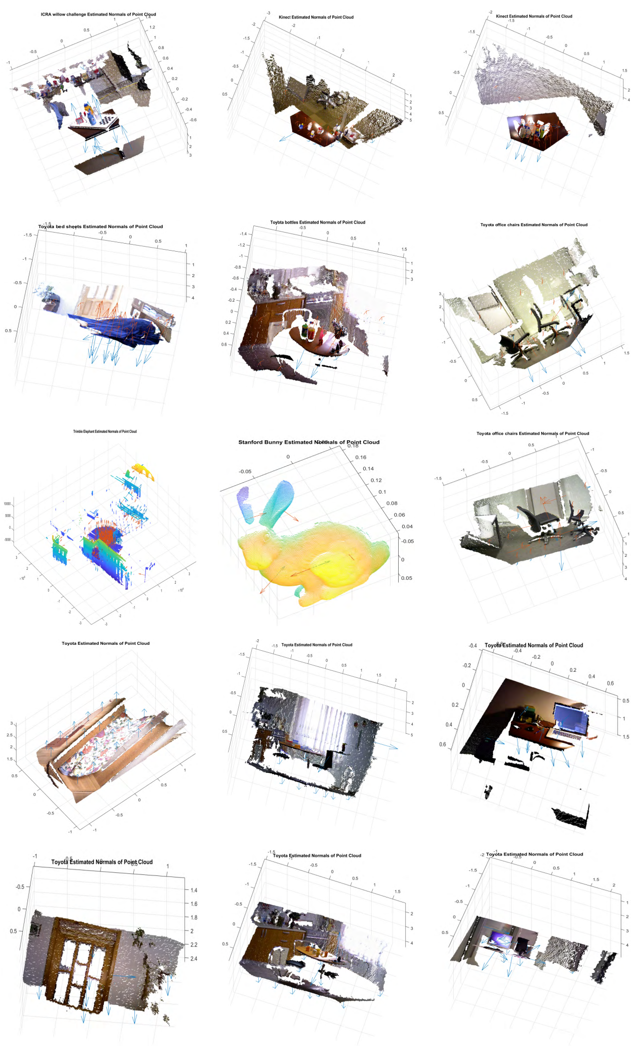


FIGURE 3. **Virtual and realistic dataset application results for the normal estimation.**

signal space is $L_{p,e}[0,\infty) \in R^N$, $1 \le p \le \infty$, which is the 3D ($N = 3$) vector signal, including the scalar signal. To simplify, $L_p(L_{p,e})$ is represented by $L_p[0,\infty)(L_{p,e}[0,\infty))$.

The signal boost of the signal, $f(t)$, is the mapping; $W_\tau : L_{p,e}[0,\infty) \to l_{L_p[0,\tau)}$,

$$\begin{aligned} \hat{f} &= W_\tau f, \\ \hat{f}_l(t) &= f(\tau i + t), \quad 0 \le t \le \tau, \quad i = 0, 1, 2 \ldots \end{aligned} \tag{3}$$

The boosting operator W_τ is regarded as cutting the continuous signal $f(t)$ into sampling signals, $\hat{f}(t)$, which are connected to each other by the sampling time τ. This sequence, $\left\{\hat{f}(t)\right\}$, is also a discrete signal, except that each element $\hat{f}(t)$ is in the function space $L_p[0,\tau)$ value.

As W_τ is a one-to-one mapping linear operator on the linear space $L_{p,e}[0,\infty)$, W_τ^{-1} is also a linear operator.

$$\begin{aligned} f &= W_\tau^{-1} g, \\ f(t) &= g_l(t - \tau i), \quad \tau i \le t \le \tau(i+1) \end{aligned} \tag{4}$$

It is an important characteristic of the boosting that W_τ is an equidistant operator, for which the norm of the signal is equal before and after the boosting. Fig. 4 indicates that the boosting of a generalized object is the promoted output y, in which the input is the inverse transform W_τ^{-1}, and the boost signal $\{\hat{u}_k\}$ is transformed into a continuous signal u, so that when the generalized object is promoted backwards and forwards, the input and output signals are the boosting signals.

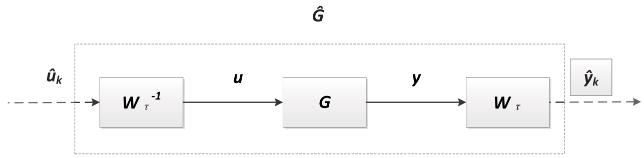


FIGURE 4. **Elevated system, $\hat{G}$.**

Because $\hat{G}$ is a linear operator, W_τ and W_τ^{-1} are isometric operators, so $\|G\| = \|\hat{G}\|$, which is the norm of the generalized object, is equal for both backward and forward boosting. Combining this with Fig. 4, the state equation of G is as follows

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \quad t \in (0,\infty) \end{aligned} \tag{5}$$

The state space of the generalized object G is implemented by

$$G = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]. \tag{6}$$

The state space of the boosting object $\hat{G}$ operator is implemented by

$$\hat{G} = \left[\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array}\right], \tag{7}$$

where

$$\begin{aligned} \hat{A} &: \mathbb{R}^x \to \mathbb{R}^x \\ \hat{B} &: L_p[0,\tau) \to \mathbb{R}^x \\ \hat{C} &: \mathbb{R}^x \to L_p[0,\tau) \\ \hat{D} &: L_p[0,\tau) \to L_p[0,\tau). \end{aligned} \tag{8}$$

For $\mathbb{R}^x$, x refers to the dimension, which is three in this paper. The discrete state space equivalent to the norm of the sampled generalized object $\tilde{G}$ is implemented as follows [26],

$$\begin{aligned} \tilde{G} &= \left[\begin{array}{c|c} \tilde{G_{11}} & \tilde{G_{12}} \\ \hline \tilde{G_{21}} & \tilde{G_{22}} \end{array}\right] = \left[\begin{array}{c|cc} \hat{A} & \hat{B_1} & \hat{B_2} \\ \hline \tilde{C_1} & \tilde{D_{11}} & \tilde{D_{12}} \\ \tilde{C_2} & 0 & 0 \end{array}\right] \\ &= \left[\begin{array}{c|cc} e^{A\tau} & e^{A(\tau-s)}B_1 & \Psi(\tau)B_2 \\ \hline C_1 e^{At} & \hat{C_1} e^{A(t-s)} l_{(t-s)} B_1 & C_1\Psi(t)B_2 + D_{12} \\ C_2 & 0 & 0 \end{array}\right], \end{aligned} \tag{9}$$

where

$$\begin{aligned} \tilde{G_{11}} &: l_{L_p[0,\tau)} \rightarrow l_{L_p[0,\tau)} \\ \tilde{G_{12}} &: l_{\mathbb{R}^u} \rightarrow l_{L_p[0,\tau)} \\ \tilde{G_{21}} &: l_{L_p[0,\tau)} \rightarrow l_{\mathbb{R}^y} \\ \tilde{G_{22}} &: l_{\mathbb{R}^u} \rightarrow l_{\mathbb{R}^y}, \end{aligned} \tag{10}$$

and u and y are the continuous signals.

H∞ discretization is the process of the generalized object promotion and equivalent discretization, and the final object is the discretized object of finite dimension equivalent to H∞ [27]. The discretization process is divided into two steps. The first step involves the use of the loop shifting method [28] to remove the pass-through term operator $\hat{D_{11}}$ from the generalized object $\tilde{G}$ in the above equation, to obtain an equivalent $\bar{G}$.

$$\bar{G} = \left[\begin{array}{c|cc} \bar{A} & \bar{B_1} & \bar{B_2} \\ \hline \bar{C_1} & 0 & \bar{D_{12}} \\ \bar{C_2} & 0 & 0 \end{array}\right]. \tag{11}$$

$\bar{G}$ is satisfied when $\|F(\bar{G}, K_d)\| < 1 \Longleftrightarrow \|F(\tilde{G}, K_d)\| < 1$, and $F(G, K)$ is the mapping between the closed-loop generalized objects from the external inputs to the regulated outputs. The second step involves the transformation of $\bar{G}$ into a finite-dimensional discretized object, which is obtained according to the following theorem [26]. If $\bar{G}$ is equivalent to the following discrete object G_d and $\|F(G_d, K_d)\| = \|F(\bar{G}, K_d)\|$, then

$$\begin{aligned} G_d &= \left[\begin{array}{c|cc} A_d & B_{1d} & B_{2d} \\ \hline C_{1d} & 0 & D_{12d} \\ C_{2d} & 0 & 0 \end{array}\right] \\ \textit{and} \quad B_{ld} &:= T_B^* \begin{bmatrix} \sum_b^{1/2} \\ 0 \end{bmatrix}, \quad \begin{bmatrix} C_{1d} & D_{12d} \end{bmatrix} := \begin{bmatrix} \sum_{cd}^{1/2} & 0 \end{bmatrix} T_C \\ A_d &:= \hat{A} + \hat{B_1}\hat{D_{11}^*}(I - \hat{D_{11}}\hat{D_{11}^*})^{-1}\hat{C_1} \\ B_{2d} &:= \hat{B_1}\hat{D_{11}^*}(I - \hat{D_{11}}\hat{D_{11}^*})^{-1}\tilde{D_{12}} + \tilde{B_2} \\ C_{2d} &:= C_2. \end{aligned} \tag{12}$$

T_B and T_{CD} are obtained according to the SVD. In fact, in H∞ control theory, the generalized object, H∞, is the peak of the maximum singular value curve in the Bode plot [29]–[32].

$$\begin{aligned} \bar{B_l}\bar{B_l^*} &= T_B^* \begin{bmatrix} \sum_b & 0 \\ 0 & 0 \end{bmatrix} T_B \\ \begin{bmatrix} \bar{C_1^*} \\ \bar{D_{12}^*} \end{bmatrix} \begin{bmatrix} \bar{C_1} & \bar{D_{12}} \end{bmatrix} &= T_{CD}^* \begin{bmatrix} \sum_{cd} & 0 \\ 0 & 0 \end{bmatrix} T_{CD} \end{aligned} \tag{13}$$

C. POSE ESTIMATION AND TRANSFORMATION

A pose transformation occurs when the coordinates of the geometric object are transformed in 3D space, which is also known as rigid object motion. The rigid object motion keeps the inner product and metric of the geometric information unchanged. The rotation in the corresponding coordinate transformation is the orthogonal matrix of the determinant one. Conversely, the rigid transformation does not have to take into account the scale factor between the set of points added during the scale transformation, compared with the isotropic or heterogeneous scale transformations that are in non-rigid transformations.

1) ROTATION MATRIX

The rotation transformation of the 3D coordinate of the point cloud is such that X, Y, and Z are the rotation axes, with corresponding rotation angles α, β and γ, respectively. The representation of the rotation matrix R is as (14) shown at the bottom of this page.

Similarly, the translation vector is set to zero, and the correspondence between the target point cloud A and reference point cloud B is as (15) shown at the bottom of this page.

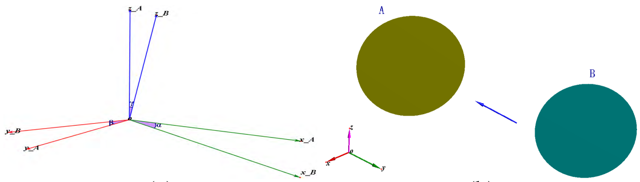
The target point cloud A is rotated only with the reference point cloud B, as shown in Fig. 5(a). B firstly rotates by angle α along the X-axis direction, then rotates by angle β along the Y-axis direction, and finally rotates by angle γ along the Z-axis direction. The rotation transformation is orderly and cannot be reversed at random.

2) TRANSLATION VECTOR

It is assumed that the 3D coordinate point cloud only needs to perform a translation transformation, when moving from left to right along the X-axis direction, translating back and forth along the Y-axis direction, and translating up-to-down along the Z-axis direction. The rotation matrix can then be represented by a matrix with main diagonal elements of one,

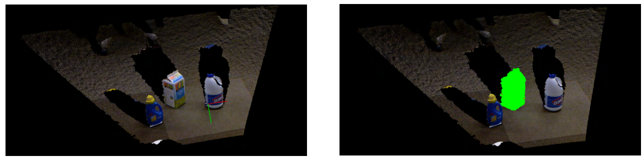
$$\begin{aligned} R &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\alpha & sin\alpha \\ 0 & -sin\alpha & cos\alpha \end{bmatrix} \begin{bmatrix} cos\beta & 0 & sin\beta \\ 0 & 1 & 0 \\ -sin\beta & 0 & cos\beta \end{bmatrix} \begin{bmatrix} cos\gamma & sin\gamma & 0 \\ -sin\gamma & cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} cos\beta cos\gamma & cos\beta sin\gamma & sin\beta \\ -cos\alpha sin\gamma - sin\alpha sin\beta cos\gamma & cos\alpha cos\gamma - sin\alpha sin\beta sin\gamma & sin\alpha cos\beta \\ sin\alpha sin\gamma - cos\alpha sin\beta cos\gamma & -sin\alpha cos\gamma - cos\alpha sin\beta sin\gamma & cos\alpha cos\beta \end{bmatrix}. \end{aligned} \tag{14}$$

$$B = \begin{bmatrix} cos\beta cos\gamma & cos\beta sin\gamma & sin\beta & 0 \\ -cos\alpha sin\gamma - sin\alpha sin\beta cos\gamma & cos\alpha cos\gamma - sin\alpha sin\beta sin\gamma & sin\alpha cos\beta & 0 \\ sin\alpha sin\gamma - cos\alpha sin\beta cos\gamma & -sin\alpha cos\gamma - cos\alpha sin\beta sin\gamma & cos\alpha cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_A \\ y_A \\ z_A \\ 1 \end{bmatrix}. \tag{15}$$



(a) (b)

FIGURE 5. Rigid object transformation. (a) Rotation transformation. (b) Translation transformation.



(a) (b)

FIGURE 6. Kinect dataset results. (a)Source scene. (b) Result scene using H∞.

with all other elements zero, which is expressed by:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{16}$$

The correspondence between the target point cloud A and reference point cloud B can then be expressed as:

$$\begin{aligned} B &= \begin{bmatrix} 1 & 0 & 0 & d_x \\ 0 & 1 & 0 & d_y \\ 0 & 0 & 1 & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \left[x_A \ y_A \ z_A \ 1\right]^T \\ &= \left[x_A + d_x \ y_A + d_y \ z_A + d_z \ 1\right]. \end{aligned} \tag{17}$$

If the target point cloud A only translates with the reference point cloud B, then Fig.5(b) indicates that B passes firstly to the left and then upwards, and finally the backward movement coincides with A.

III. EXPERIMENTAL RESULTS

For the experiment, the proposed method was implemented on the Visual Studio 2013 platform, in the Point Cloud Library (PCL) [33]. The programming language was C++. The Kinect, Mian [34], and Clutter [35] datasets were used. The detailed configuration project parameters of the H∞ optimal control in the C++ project are given in Table 2. The functional description of the various parameters are also described. There are only a few PCL configuration parameters because the exact value is automatically obtained by H∞ and Jordan SVD. The first experiment is based on the Kinect dataset as shown in Fig. 6, which is the source scene pictures. As shown in Fig. 6, after the posture estimation, the shape of the milk box is unchanged, position is accurate, outline of the edge of the box is moderate, and thickness is uniform. Therefore, the effect of the proposed method is satisfactory.

The second experiment was based on the Mian dataset. There are five models and fifty scenes in this dataset. As before, the pose estimation is the vertex position coordinates of the model estimation, which is the display correction of the model. The model and scene were used simultaneously to correct the experiment to avoid confusing the same model as both input and output. Moreover, by performing simultaneous operations of the model and scene, the vertex position coordinates can be obtained more accurately. Since the principle of H∞ optimal control is based on the maximum singular value of the rational function matrix parsed in the right half plane, the first step is to calculate the right singular vector. Fig. 7 presents the sequential matrix value plots of five models for the right vector of fifty scenes in the Mian dataset. There are twenty-five in each row, and two rows represent a model. The names of the models are chef, chicken_high, parasaurolophus_high, rhino, and T-rex_high. Fig. 8 shows the comparison of the maximum singular value to and fro of the H∞ optimal control. According to the definition of H∞, the maximum singular value is the best result. This can be observed in Fig. 8 in which without H∞ optimal control, the maximum singular value is random and not sufficiently accurate. Figs. 9(a) and 9(b) show the results with and without H∞ control, respectively. As shown in Fig. 9(a), the incorporating and matching effects between the model and scene are good. The hat of the chef fits well, and the coordinates

TABLE 2. Project parameters of the H∞ optimal control.

Symbol	Value[Types]	Description
leaf	0.005[float]	Point cloud voxel grid size
setRadiusSearch	0.01[double]	Normal estimation search radius
setFeaturesRadiusSearch	0.025[double]	Normal features estimation search radius
setMaximumIteration	50000[int]	Number of RANSAC iterations
setNumberOfSamples	3[int]	Number of points to sample for generating or prerejecting a pose
setCorrespondenceRandomness	65[int]	Number of nearest features to use
setSimilarityThreshold	0.9[float]	Polygonal edge length similarity threshold
setMaxCorrespondenceDistance	2.5[float] × leaf	Inlier threshold
setInlierFraction	0.25[float]	Required inlier fraction for accepting a pose hypothesis

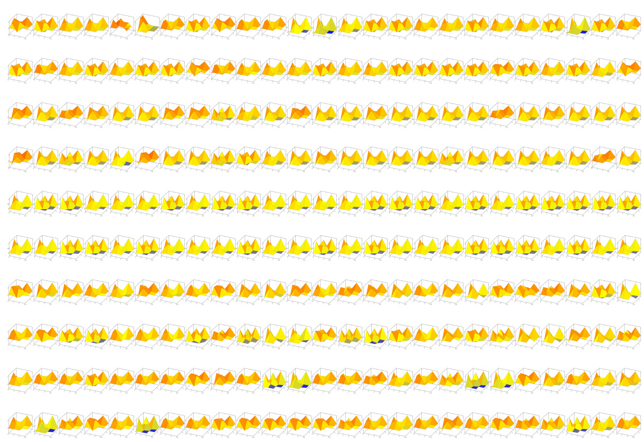


FIGURE 7. Experimental results of the right singular vector for the Mian dataset.

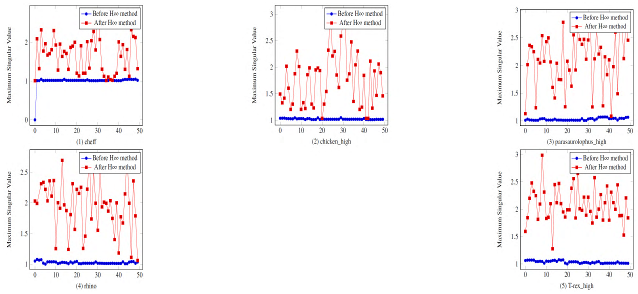


FIGURE 8. Comparison experimental results of the Mian dataset.

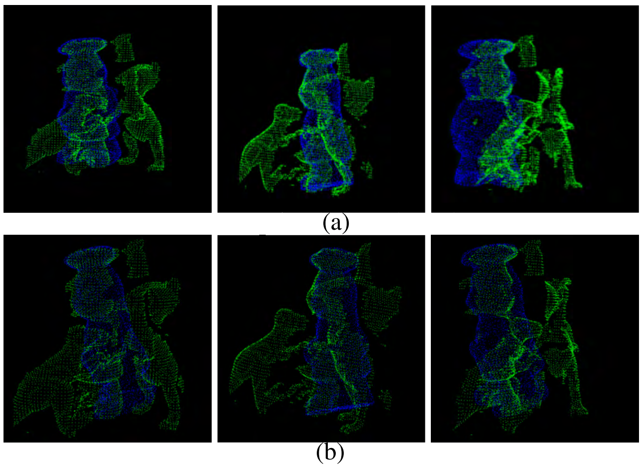


FIGURE 9. Experimental rendering of the Mian dataset. (a) H∞ optimal control results. (b) Results without H∞.

and matching of his face, especially the ear, and his body position are also accurate, therefore the position estimation of a rigid object had a robustness effect, as shown in the pictures. However, Fig. 9(b), which is the picture without H∞ control and was performed under the current hardware platform, reveals that the pose estimation of the object was inaccurate, and therefore, subsequent manual adjustments are required. In short, after comparison, the proposed method is an effective approach for pose correction.

The third experiment was based on the Clutter dataset. There are eighteen models and thirty scenes in this dataset. Fig. 10 shows the matrix value plot of eighteen models for the right vector of thirty scenes in the clutter dataset. There are thirty in each row, and one row represents a model. The names of the models were 409Bottle, BakingSodaBox, Banana, BlueBowl, CascadeBottle, GreenBrush, GreenPear, HootBot, JiffyBox, LBlock, MetalMug, MongoDBMug, PaperCup, RectCup, StrawBowl, TikiCup, TriangleBox, and Yellow-Pepper. Fig. 11 shows the comparison of the maximum singular value to and fro of the H∞ optimal control. As shown in Fig. 11, the results of the proposed method are better than those of the method without using H∞. This reflects the superiority of the propsed method. Figs. 12(a) and 12(b) also show the results with and without H∞ control, respectively. As shown in Fig. 12, Fig. 12(a) demonstrates the robustness of the pose estimation through the model and scene.

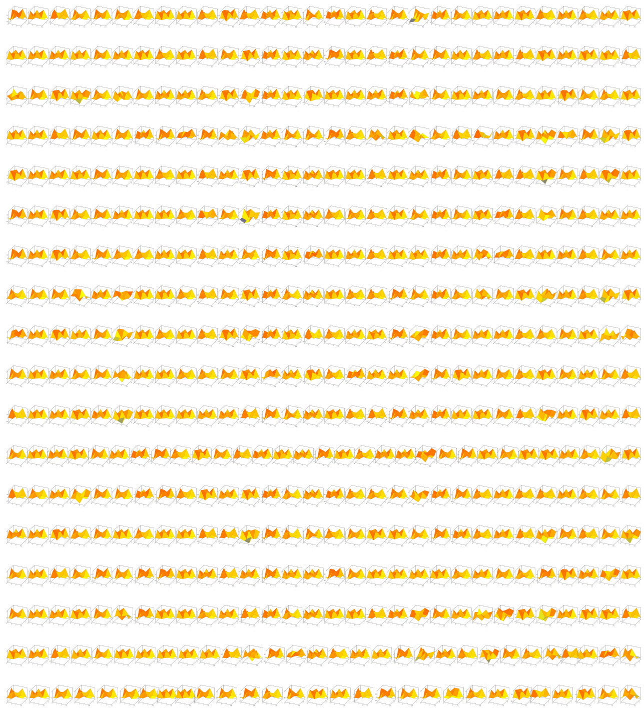


FIGURE 10. Experimental results of the right singular vector for the Clutter dataset.

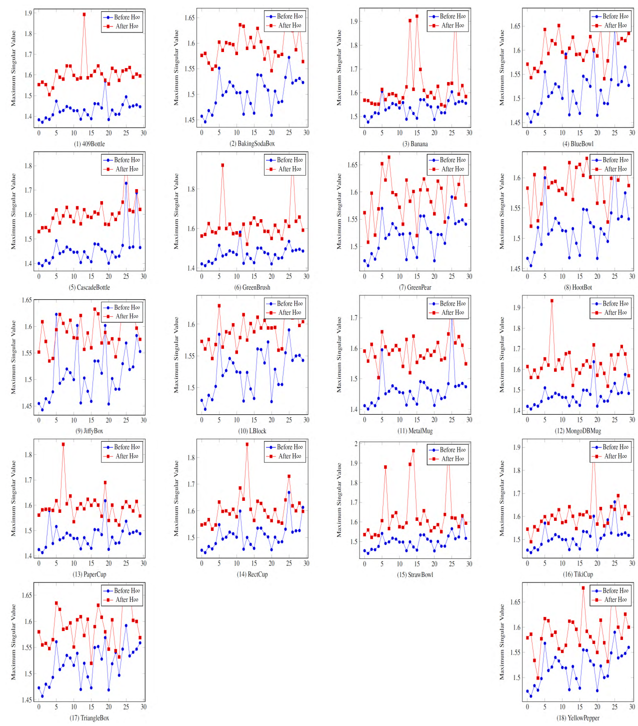


FIGURE 11. Comparison experimental results of the Clutter dataset.

From Fig. 12(b), the pose estimation deviated without H∞ control under the current hardware platform. Overall, after the comparison, the proposed approach is still an effective method.

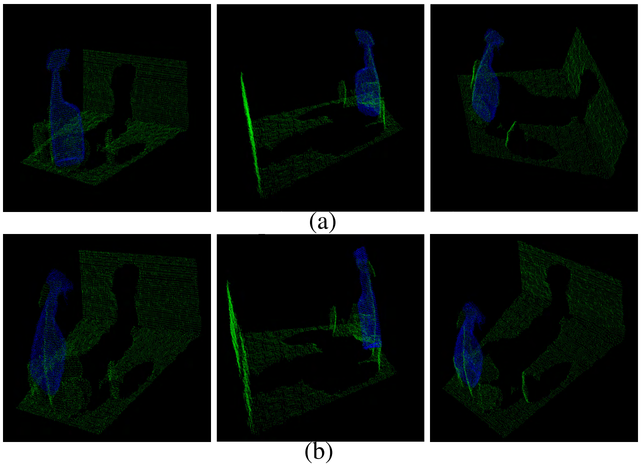


FIGURE 12. Experimental rendering of the Clutter dataset. (a) H∞ optimal control results. (b) Results without H ∞.

IV. CONCLUSION

At present, the popular light field 3D display technology is a better solution for naked eye 3D displays. In this paper, we proposed a novel H∞ method that does not require the addition of hardware or software, which from a mathematical point of view, is to bring about the position correction. The experimental results showed that the proposed approach is robust.

In summary, we proposed the H∞ optimal control method in the light field of a 3D display. In addition, we performed many experimental and theoretical studies of the new model. The experimental results support the validity of the proposed approach, with a high accuracy of the maximum singular value and a robust pose correction, compared with the current pure hardware platform method.

In the future, we carry on to add the Gaussian mapping and K-means clustering in the normal estimation module, which is a new innovation of the preprocessing stage before the H∞ method in the 3D light field imaging.

REFERENCES

[1] H. Shidanshidi, F. Safeai, and W. Li, "Estimation of signal distortion using effective sampling density for light field-based free viewpoint video," *IEEE Trans. Multimedi*, vol. 17, no. 10, pp. 1677–1693, Oct. 2015.

[2] C. Conti, L. D. Soares, and P. Nunes, "Light field coding with field-of-view scalability and exemplar-based interlayer prediction," *IEEE Trans. Multimedi*, vol. 20, no. 11, pp. 2905–2920, 2018.

[3] Y. Peng, "Self-calibration three-dimensional light field display based on scalable multi-LCDs," *Soc. Inf. Display*, vol. 20, no. 12, pp. 653–660, Sep. 2012.

[4] Y. Peng, "Large-sized light field three-dimensional display using multi-projectors and directional diffuser," *Opt. Eng.*, vol. 52, no. 1, Jan. 2013, Art. no. 017402.

[5] Q. Zhong, Y. Peng, H. Li, and X. Liu, "Optimized image synthesis for multi-projector-type light field display," *Display Technol.*, vol. 12, no. 12, pp. 1745–1751, Dec. 2016.

[6] T. Zhan, "High-resolution additive light field near-eye display by switchable Pancharatnam-Berry phase lenses," *Opt. Express*, vol. 26, no. 4, pp. 4863–4872, 2018.

[7] M. Halle, "Multiple viewpoint rendering," in *Proc. ACM SIGGRAPH*, Jul. 1998, pp. 242–254.

[8] X. Hou, "Real-time multi-perspective rendering on graphics hardware," in *Proc. 17th Eurographics Conf. Rendering Techn.*, Aug. 2006, pp. 93–102.

[9] A. Jones, "Rendering for an interactive 360° light field display," *ACM Trans. Graph.*, vol. 26, no. 3, 2007, Art. no. 40.

[10] A. Butler, "Vermeer: Direct interaction with a 360° viewable 3D display," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, Santa Barbara, CA, USA, Sep. 2011, pp. 16–19.

[11] X. Xia, "A 360-degree floating 3D display based on light field regeneration," *J. Opt. Express*, vol. 21, no. 9, pp. 11237–11247, 2013.

[12] J. Navarro and A. Buades, "Reliable light field multiwindow disparity estimation," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 1449–1453.

[13] J. Navarro and A. Buades, "Robust and dense depth estimation for light field images," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1873–1886, Apr. 2017.

[14] J.-W. Wang, N. T. Le, C. C. Wang, and J. S. Lee, "Hand posture recognition using a three-dimensional light field camera," *IEEE Sensors J.*, vol. 16, no. 11, pp. 4389–4396, Jun. 2016.

[15] C. Jia, F. Shi, Y. Zhao, Z. Wang, M. Zhao, and S. Chen, "Identification of pedestrians from confused planar objects using light field imaging," *IEEE Access*, vol. 6, pp. 39375–39384, 2018.

[16] X. Huang, Z. Huang, M. Lu, P. Ma, and W. Ding, "A semi-global matching method for large-scale light field images," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1646–1650.

[17] J. Santamaría, M. A. Marquez, and S. Pertuz, "A method for improving depth estimation in light field images," in *Proc. 11th Symp. Signal Process., Images Artif. Vis. (STSIVA)*. Bucaramanga, CA, USA, Aug. 2016, pp. 1–7.

[18] R. Ferreira and N. Gonçalves, "Accurate and fast micro lenses depth maps from a 3D point cloud in light field cameras," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*. Dec. 2016, pp. 1893–1898.

[19] Y. Zhang *et al.*, "Light-field depth estimation via epipolar plane image analysis and locally linear embedding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 4, pp. 739–747, Apr. 2017.

[20] O. Johannsen *et al.*, "A taxonomy and evaluation of dense light field depth estimation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1795–1812.

[21] P. de Oliveira Rente, C. Brites, J. Ascenso, and F. Pereira, "Graph-based static 3D point clouds geometry coding," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 284–299, Feb. 2019.

[22] Y. Anisimov and D. Stricker, "Fast and efficient depth map estimation from light fields," in *Proc. Int. Conf. 3D Vis.*, Sep. 2017, pp. 337–346.

[23] J. Yu, "A light-field journey to virtual reality," *IEEE MultiMedia*, vol. 24, no. 2, pp. 104–112, Apr. 2017.

[24] G. Wu *et al.*, "Light field image processing: An overview," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, pp. 926–954, Oct. 2017.

[25] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *ACM SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 71–78, 1992.

[26] B. A. Bamieh and J. Pearson, Jr., "A general framework for linear periodic systems with applications to $H^{infinity}$ sampled-data control," *IEEE Trans. Autom. Control*, vol. 37, no. 4, pp. 418–435, Apr. 1992.

[27] L. Mirkin and G. Tadmor, "Yet another H^{∞} discretization," *IEEE Trans. Autom. Control*, vol. 48, no. 5, pp. 891–894, May 2003.

[28] M. G. Safonov and D. J. N. Limebeer, "Simplifying the H∞ theory via loop shifting," in *Proc. 27th IEEE Conf. Decision Control*, Dec. 1988, pp. 1399–1404.

[29] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Trans. Autom. Control*, vol. 26, no. 2, pp. 301–320, Apr. 1981.

[30] J. C. Doyle, "State space solution to standard H2 and H∞ control problem," *IEEE Trans. Autom. Control*, vol. 34, no. 8, pp. 831–847, Aug. 1989.

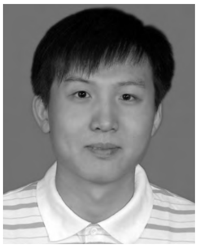
[31] T. Chen, "H∞-optimal sampled-data control computation and design," *Elsevier Sci. Automatica*, vol. 32, no. 2, pp. 223–228, 1996.

[32] R. W. Beard, "Linear operator equations with applications in control and signal processing," *IEEE Control Syst.*, vol. 22, no. 2, pp. 69–79, Apr. 2002.

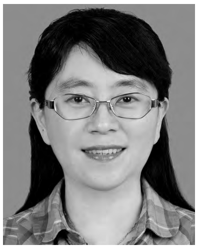
[33] Point Cloud Library(PCL) 1.8.0. (2017). *Point Cloud Library: The Standalone, Large Scale, Open Project for 2D/3D Image and Point Cloud Processing*. [Online]. Available: http://www.pointclouds.org/

[34] A. Mian, "3D model-based object recognition and segmentation in cluttered scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1584–1601, Oct. 2006.

[35] J. Glover and S. Popovic, "Bingham procrustean alignment for object detection in clutter," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Aug. 2013, pp. 2158–2165.



ZHENHAO WANG received the B.S. degree in electronic information engineering from the Changchun University of Technology, in 2006, and the M.S. degree in electronic science technology from the Changchun University of Science and Technology, in 2012. He is currently pursuing the Ph.D. degree in communication and information systems with Jilin University. He has worked in (two- and three-dimensional) image processing algorithms for nine years. His current researches relate to three-dimensional image processing algorithm design, and associated applications in light field imaging and augmented reality.



YAN ZHAO received the B.S. degree in communication engineering from the Changchun Institute of Posts and Telecommunications, in 1993, the M.S. degree in communications and electronics from the Jilin University of Technology, in 1999, and the Ph.D. degree in communication and information systems from Jilin University, in 2003, where she is currently a Professor of communication engineering. Her research interest includes image and video processing.

SHIGANG WANG received the B.S. degree from Northeastern University, in 1983, the M.S degree in communication and electronics from the Jilin University of Technology, in 1998, and the Ph.D. degree in communication and information systems from Jilin University, in 2001, where he is currently a Professor of communication engineering. His research interests include multidimensional signal processing, and stereoscopic and multi-view video coding.