IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 17, 2019, accepted May 28, 2019, date of publication June 4, 2019, date of current version June 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920278

Mixed Probability Inverse Depth Estimation Based on Probabilistic Graph Model

WENLEI LIU[1], SENTANG WU[1], XIAOLONG WU[2], AND KAI LI[1]
[1]School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China
[2]Navigation and Control Technology Institute of NORINCO Group, Beijing 100089, China

Corresponding author: Wenlei Liu (liuwenlei@ buaa.edu.cn)

This work was supported by the Industrial Technology Development Program under Grant B1120131046.

ABSTRACT In this paper, a mixed probability inverse depth estimation method based on probabilistic graph model is proposed, which can effectively solve the problems of far distance from the camera center and long data tail in depth estimation. At the same time, not only the accuracy can be improved but also the robustness of inverse depth estimation can be developed. First, the triangle method was used to find the depth information and location of a point in space, and the inverse depth information was obtained as the initial information of inverse depth estimation. Then, the basic matrix in epipolar geometry was obtained by using the normalized eight-point algorithm, and the pose of a camera was obtained as the initial information of optimization. Next, the pose of the monocular camera was modeled by a factor graph model, and the pose estimation was transformed into an unconstrained optimization problem by using the transformation relationship between Lie group and Lie algebra to obtain the pose of the camera. Finally, the inverse depth obtained by using the Gauss-uniform mixed probability distribution based on the probability graph model was used to calculate the recurrence formula by approximate inference, which can facilitate the sequential processing of multiple images. The depth information was quantitatively measured and compared by using TUM datasets, and the length of space object was measured by using inverse depth information, thus the measurement accuracy of this method was indirectly verified. This method is characterized by strong robustness and high measurement accuracy in the environments with random interferences.

INDEX TERMS Mixed probability distribution model, factor graph, inverse depth estimation, Lie group and Lie algebra, fundamental matrix, camera pose.

I. INTRODUCTION

With the development of the artificial intelligence technology, research on SLAM of monocular vision goes further and further. One of the important research field is the estimation of depth information of monocular vision, which is also a difficult point of monocular vision research. The accuracy of SLAM location can be guaranteed only if the depth information is estimated precisely. In order to effectively solve the depth estimation problem of monocular vision SLAM, machine learning method is applied to depth estimation, so a model of Gauss-Uniform mixture probability distribution is considered to improve the robustness and accuracy of the system. In order to solve the data tailing phenomenon in depth estimation, the concept of inverse depth is introduced into estimation. The inverse depth estimation based on Gauss-Uniform Mixed Probability Distribution can not only solve the problem of data tailing, but also effectively improve the measurement accuracy and robustness.

As an important research field in monocular vision, there are many studies having been conducted on depth estimation. In reference [1], proposed was a method of depth map calculation based on Bayesian estimation and convex optimization image processing. This method can save memory and improve computing speed. In reference [2], a real-time density tracking and mapping method was proposed to estimate detailed texture depth mapping on selected key frames to generate surface mosaics containing millions of vertices. Reference [3] proposed a semi-direct monocular vision ranging algorithm, which uses an explicit probability mapping method to model outliers and to estimate three-dimensional points, so as to obtain fewer outliers and more reliable points.

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

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

In reference [4], in order to improve the accuracy of camera estimation of depth and spatial position, different sensors were used to register a three-dimensional map calculated by a robot, to improve its accuracy. Reference [5] proposed an inverse depth parameterization method based on traceless transformation, which can capture the relationship between low parallax features and error state variables of unknown depth information. Reference [6] proposed a fast method to solve camera depth by using distance flow constraint equation and sparse geometric features, and applies its depth estimation to visual SLAM. In reference [7], Bayesian estimation method was applied to depth estimation. The above methods all are based on the assumption that the depth of a point agrees with the Gauss distribution, but there are data tailing at the position of the principle camera, which may cause great errors in measurement.

In order to solve the problem of data tailing in depth estimation, the inverse depth method is devised to meet this currently inevitable requirement. The inverse depth not only makes the data distribution more reasonable, but also improves the estimation accuracy. References [8], [9] proposed a monocular vision SLAM based on inverse depth parameterization, the method of inverse depth parameterization which has been introduced in details. The parameterization for this method can cope with features in a large depth range. The features far from the camera, even those located at infinite distance, still maintain enough representativeness in the course of motion, with almost no parallax. The inverse depth of direct parameterization causes a problem of increasing number of parameters, which brings some difficulties to optimization. In the process of optimization, the constraints between parameters are not considered. Reference [10] proposed smoothing and mapping (SAM) be applied to inverse depth parameter estimation, while avoiding overgeneralization of parameters. Results with excellent accuracy were obtained from real data. The semi-direct visual localization (SDVL) method was applied for inverse depth estimation, which improves the efficiency of feature matching, and the anomalous rejection mechanism (ORM) was used to eliminate the dislocation [11]. Although inverse depth's parametric estimation improves the accuracy of parameters, the constraints between them are usually neglected in order to avoid over-parameterization.

By the transformation between Lie group and Lie algebra, the estimation problem can be transformed into an unconstrained optimization problem, to simplify the solution of the problem. References [12]–[14] described how to use Lie group and Lie algebra to represent rotation and pose in three-dimensional space, how to transform a constrained rotation matrix into an unconstrained optimization problem, and how to derive the Jacobian matrix. The Lie group-Lie algebra is a mathematical tool to simplify the problem of attitude derivation, which brings convenient solutions.

It depends largely on its pose estimation accuracy how accurate the depth estimation of monocular camera is. Therefore, pose estimation plays an important role in depth estimation. In reference [15], a pose map was used to optimize the camera attitude. This method is widely used in low dynamic environment, which reduces the computational load and improves the running speed. Reference [16] proposed a closed-loop online pose chain method to cope with pose estimation, which can accurately estimate the depth of visual odometer and reduce the influence of scale drift on pose estimation. In reference [17], the factor graph was applied to the complex pose estimation problem. This method is applicable for the case of random position interference in the estimation. Using probability information, unknown variables can be predicted. An open-square smoothing filtering algorithm was proposed to estimate the pose of the camera, which decomposes correlation information matrix or measured Jacobian matrix into square roots, so the calculation speed is faster, and the accuracy is higher [18]. And a new data structure, namely Bayesian tree algorithm, has been proposed. The method of probability graph was used to infer pose, and sparse matrix decomposition was used to improve the speed and accuracy of operation [19]. References [20]–[22] used incremental smoothing and mapping to decompose information matrix, which only updates the actual matrix items, thus improving the efficiency, and used uncertainty estimation algorithm based on factor information matrix to improve the real-time information processing.

Random interference in experimental processes could have a serious impact on the results. In order to improve the robustness of the estimation, many researches have been done. In references [23], [24], a pixel-by-pixel probabilistic depth estimation scheme was proposed. Posterior depth distribution can be updated with update of each frame. The Gauss-Uniform mixture probability distribution pattern was adopted to improve the robustness of the system. The maximum likelihood estimation method based on non-Uniform mixture parameters with Gauss-Uniform distribution was used to solve the estimation problem under constraints, and the existence and consistency of the method were proved [25]. Reference [26] introduced a mixed model of multivariate Gauss distribution and multivariate Uniform distribution, and applied it to clustering and classification of models, in which the validity of the mixed model was shown. In reference [27], a linear system identification method based on abnormal robust regular kernel was proposed. Unknown variables were modeled into Gauss processes, and noise signals were modeled into Laplace random variables, which improves the robustness of the system.

Many researchers also consider improving the accuracy and speed of depth estimation from the aspect of hardware. Reference [28] proposed a hardware architecture for motion depth estimation, which includes a depth conversion and a new optical flow algorithm, i.e., a pixel-parallel/window-parallel method for computing optical flow based on correlation function of absolute difference sum. In reference [29], a reactive obstacle avoidance system was proposed, which uses online adaptive convolution neural

network (CNN) to improve the depth estimation of monocular camera in unfamiliar environment step by step, and uses motion stereo image as training data. In reference [30], a system for estimating depth using tilted lens optics was proposed. The method only obtains the depth of each pixel from the sharpness ratio of two tilted optical images, and uses the method of neural network to improve the accuracy.

The main contributions of this paper include: the inverse depth estimation method was used to solve the problem of abnormal Gaussian distribution in depth estimation, which improves the accuracy and stability of the estimation; the inverse depth estimation problem was modeled by factor graph model, which facilitates the extension of the inverse depth estimation method; the transformation relationship between Lie group and Lie algebra was used to transform the solution of pose into unconstrained optimization problem, which simplifies the solution process; the inverse depth estimation method based on the Gauss-Uniform Mixture Probability Distribution Model was used; the recurrence formula was derived by approximate inference, which can facilitate sequential processing of multiple images, thus not only improving the accuracy of estimation, but also improving the robustness of the system.

The organizational structure of this paper is as follows: Section II will mainly introduce the related definitions and basic principles of this paper; in Section III will introduce the use of factor graph model to optimize camera pose, and how to transform the problem into an unconstrained problem by using the transformation between Lie group and Lie algebra; Section IV will introduce the method of estimating inverse depth based on Gauss-Uniform Mixed Probability Distribution of Probability Graph; Section VI will introduce the related experiments in indoor and outdoor environments, and analyze the experiment results; Section ćö will summarize the methods provided in this paper.

II. BASIC DEFINITIONS AND RELATED PRINCIPLES

A. USING TRIANGLE METHOD TO FIND DEPTH AND LOCATION INFORMATION OF SPACE POINT

(a) Obtaining the depth information

By observing the angle at the same point in two places, the distance of the point can be determined.

$$s_1 x = KX, \quad s_2 x' = K(RX + t) \tag{1}$$

Let $y = K^{-1}x$, $y' = K^{-1}x'$, then the following equation can be obtained from formula (1):

$$s_2 y' = s_1 R y + t \tag{2}$$

$$s_2 y' \times y' = s_1 y' \times R y + y' \times t = 0 \tag{3}$$

Depth information s_1 and s_2 can be obtained by using (2) (3).

(b) Obtaining the spatial location information

$$\begin{cases} s_1 x = PX = K[R|t]X \\ s_2 x' = P'X = K[R'|t']X \end{cases} \tag{4}$$

$$\begin{cases} x \times PX = 0 \\ x' \times P'X = 0 \end{cases} \tag{5}$$

$$\begin{bmatrix} x_1 P^{3T} - P^{1T} \\ y_1 P^{3T} - P^{2T} \\ x_2 P'^{3T} - P'^{1T} \\ y_2 P'^{3T} - P'^{2T} \end{bmatrix} \cdot X = A \cdot X = 0 \tag{6}$$

where, P^{iT} is the line of P.

X can be obtained by using the method of non-linear least squares. Because of influences of noise, it may not strictly equal to 0. The points $\bar{x'}, \bar{x}$ near them should accurately satisfy the geometric constraints $\bar{x'}^T F \bar{x} = 0$. The minimum cost function for seeking $\bar{x'}, \bar{x}$ is:

$$\begin{cases} L(x, x') = d(x, \bar{x})^2 + d(x', \bar{x'})^2 \\ \bar{x'}^T F \bar{x} = 0 \end{cases} \tag{7}$$

Under the assumption of Gauss error distribution, $\bar{x'}, \bar{x}$ is the maximum likelihood estimation of corresponding points of the real image. Once $\bar{x'}, \bar{x}$ is obtained, the space point X can be obtained by triangle method.

In general, the first order geometric correction (Sampson approximation) is usually used to find the ideal point pair $\bar{x'}, \bar{x}$; as the measurement point $Y = (x_1, y_1, x_2, y_2)^T$, the δ_x corrected by Sampson is as follows:

$$\delta_x = -J^T (JJ^T)^{-1} \varepsilon \tag{8}$$

Among them, error $\varepsilon = x'^T F x$, and Jacobian matrix is:

$$J = \partial\varepsilon / \partial x = \left[(F^T x')_1, (F^T x')_2, (Fx)_1, (Fx)_2 \right] \tag{9}$$

where, $(Fx)_i$ is the partial derivative of the ith variable.

$$\bar{Y} = Y + \delta_x = Y - J^T (JJ^T)^{-1} \varepsilon \tag{10}$$

$$\begin{bmatrix} \bar{x}_1 \\ \bar{y}_1 \\ \bar{x}_2 \\ \bar{y}_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix} - \frac{x'^T F x}{(F^T x')_1^2 + (F^T x')_2^2 + (Fx)_1^2 + (Fx)_2^2} \times \begin{bmatrix} (F^T x')_1 \\ (F^T x')_2 \\ (Fx)_1 \\ (Fx)_2 \end{bmatrix} \tag{11}$$

According to the above formula, the ideal point $\bar{x'}, \bar{x}$ can be obtained, and then X can be obtained.

Triangular measurement is mainly obtained by translation. The triangle method can be used in all cases other than rotation. The larger the translation, the higher the measurement accuracy will be. There are usually two ways to improve the accuracy of the triangle method: 1. to improve the extraction accuracy of feature points, that is, to improve the resolution, which will increase the calculation cost; 2. to increase the translation distance. With the triangle method, only the depth and location of a feature point can be measured, but the relationship between global information and feature points cannot be obtained.

B. SOLUTION OF INITIAL POSTURE

The basic matrix F is a transfer mapping from one image to another through arbitrary plane π. Assuming that the two images are obtained by a camera whose center does not coincide, both the corresponding points x and x' of the basic matrix F satisfy: [31]

$$x'^T F x = 0 \tag{12}$$

where, F is a 3×3 homogeneous matrix with a rank of 2.

Essential matrix E is a special form of fundamental matrix in normalized image coordinates. The relationship between fundamental matrix and essential matrix is as follows:

$$E = K'^T F K \tag{13}$$

A matrix is an essential matrix if and only if two of its singular values are equal and the third is 0.

The steps to solve the pose are as follows:

(a) The basic matrix F is obtained by using the classical eight-point algorithm, and the essential matrix E is obtained by formula (13).

(b) The essential matrix E is decomposed by SVD: $E = UDV^T$, in which $D = diag(\sigma_1, \sigma_2, \sigma_3)$ and $\sigma_1 \geq \sigma_2 \geq \sigma_3$. The essential matrix closest to E under the Frobenius norm is $\tilde{E} = U\tilde{D}V^T$, where $\tilde{D} = diag((\sigma_1 + \sigma_2)/2, (\sigma_1 + \sigma_2)/2, 0)$. Therefore, the essential matrix E can be simplified to $E = U \sum V^T$, where $\sum = diag(1, 1, 0)$ and $U, V \in SO(3)$.

(c) Recovering the pose of the camera from the essential matrix: Assuming that the first camera is the coordinate origin, i.e. $P = [I|0]$, according to the essential matrix E expressed as $E = Udiag(1, 1, 0)V^T$, the R and t of the second camera can be obtained as follows [13]:

$$R = UR_Z^T(\pm\frac{\pi}{2})V^T, \quad t = UR_Z(\pm\frac{\pi}{2})\sum U^T \tag{14}$$

where, $R_Z(\theta)$ represents the rotation matrix of θ degree around Z axis (or X_3 axis).

According to the result of the third step, there are four groups of solutions, but only one group of them has a positive image depth. Therefore, we choose a test point in the image and use the four groups to solve the depth of the point. The group with a positive depth is the final solution.

C. INVERSE DEPTH FILTER

Depth estimation is usually applied to indoor application scenarios with limited distances. In complex outdoor environments, distances are very far or even infinite. In this case, the assumption of normal distribution is not valid, the negative region is zero, and the tail may be a little longer. In order to solve the problem of remote depth estimation, the idea of inverse depth comes into being. In practical applications, the inverse depth has better numerical stability.

An inverse depth filter is used to make the inverse depth estimation converge gradually from an uncertain inverse depth value to a stable value along with the increase of measured data, and to estimate the change of the inverse depth distribution. Usually it comprises epipolar search and block matching technology. For a point in an image, its inverse depth is unknown, and the corresponding spatial points are distributed in a line segment. From another perspective, the projection of this line segment is a line of the image plane, which is an epipolar line. Block matching technology is to select a $w \times w$ block near the matching point and compare the same size blocks on the epipolar line, which can improve the resolution to a certain extent. Assuming that the blocks of matching points are represented by A_i and the blocks of epipolar lines are represented by B_i, there are three evaluation methods available:

(a) SAD: the sum of absolute values of difference between two small blocks;

$$S(A, B)_{SAD} = \sum_{i,j} |A(i, j) - B(i, j)| \tag{15}$$

(b) SSD: the sum of squares of differences;

$$S(A, B)_{SSD} = \sum_{i,j} (A(i, j) - B(i, j))^2 \tag{16}$$

(c) NCC: Normalized cross correlation, in which the correlation between two small blocks needs to be calculated.

$$S(A, B)_{NCC} = \frac{\sum_{i,j} A(i, j)B(i, j)}{\sqrt{\sum_{i,j} A(i, j)^2 \sum_{i,j} B(i, j)^2}} \tag{17}$$

This paper used the NCC method in evaluation.

Assuming that the inverse depth ρ of the pixels obeys the Gauss distribution:

$$P(\rho) = N(\mu, \sigma^2) \tag{18}$$

The newly observed inverse depth data also obey the Gauss distribution:

$$P(\rho') = N(\mu', \sigma'^2) \tag{19}$$

The original depth information and observed data integrated together still complies with the Gauss distribution.

$$P(\tilde{d}) = N(\tilde{\mu}, \tilde{\sigma}^2), \quad \tilde{\mu} = \frac{\sigma'^2\mu + \sigma^2\mu'}{\sigma^2 + \sigma'^2}, \ \tilde{\sigma}^2 = \frac{\sigma^2\sigma'^2}{\sigma^2 + \sigma'^2} \tag{20}$$

Considering only the geometric uncertainties, the calculation of μ', σ'^2 is as follows by using the basic principles and properties of polar geometry [31].

The inverse depth uncertainty relationship is shown in Fig. 1. Supposing $\overrightarrow{CC'}$ is a, $\overrightarrow{CX}$ is d, $\overrightarrow{C'X}$ is c, $\overrightarrow{CX'}$ is d', as shown in Fig. 1, the two bottom angles are α, β and the top angle is γ. Supposing there is an error of pixel size on the epipolar line l_2, which makes β become β', ρ become ρ', and γ become γ', we can solve for d by triangle method. Disturbing x_2 by a pixel will make β produce a change $\delta\beta$.Since the focal length of the camera is f, and τ is the size of each pixel, it can be obtained that:

$$\delta\beta = \arctan\frac{1 * \tau}{f} \tag{21}$$

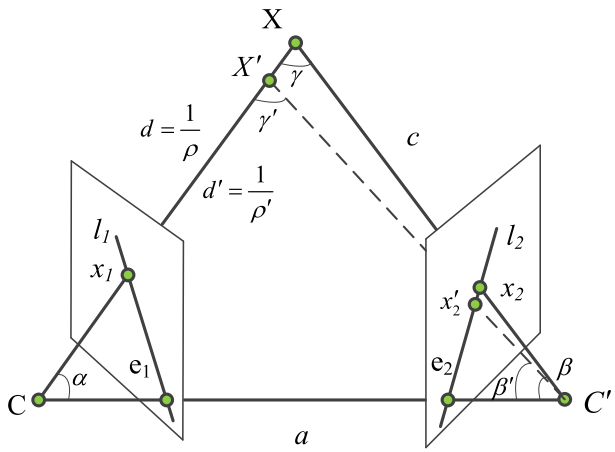

FIGURE 1. Relation analysis diagram of inverse depth uncertainty.

According to the geometric relation, we can get:

$$\begin{cases} \beta' = \beta + \delta\beta \\ \gamma' = \pi - \alpha - \beta' \end{cases} \tag{22}$$

According to the sine theorem, the size of $d^{'}$ can be obtained as follows:

$$\left\|d'\right\| = \|a\| \frac{\sin\beta'}{\sin\gamma'} \tag{23}$$

Namely, $\mu' = 1/\left\|d'\right\|$ Therefore, σ'^2 can be estimated as:

$$\sigma'^2 = \left(1/\|d\| - 1/\left\|d'\right\|\right)^2 \tag{24}$$

When the uncertainty is below a certain threshold, it can be considered that the inverse depth data has converged.

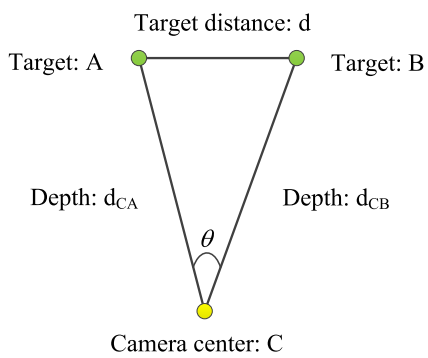

FIGURE 2. Principle of ranging based on camera depth information.

D. RANGING METHOD TO USE DEPTH INFORMATION

Fig. 2 is a schematic diagram for estimating the distance between two points in an image using the depth information measured by a monocular camera. Where, θ is the angle between the camera center and two target points. Usually, the direction vector of depth information can be used to find the angle: $\cos\theta = \frac{d_{CA} \cdot d_{CB}}{\|d_{CA}\| * \|d_{CB}\|}$.

Given the depth information d_{CA}, d_{CB} of the two target points and the angle θ between them and the camera center, the distance between the two points can be obtained by using the cosine theorem.

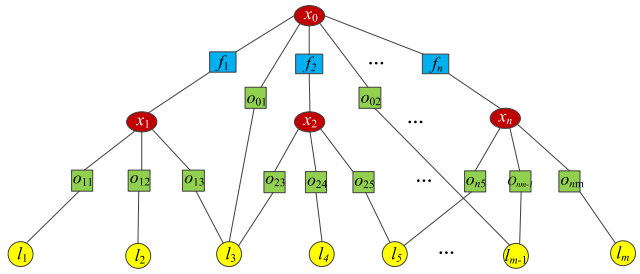

FIGURE 3. Factor diagram model diagram.

$$d^2 = d_{CA}^2 + d_{CB}^2 - 2d_{CA}d_{CB}\cos\theta \tag{25}$$

E. STEPS FOR MIXED PROBABILITY INVERSE DEPTH ESTIMATION BASED ON PROBABILITY GRAPH MODEL

The detailed implementation steps of the mixed probability inverse depth estimation method based on the probabilistic graph model proposed in this paper are as follows:

a. The classical triangle method is used to find the depth information and position of a point in space, and the first-order geometric correction method is used to improve the accuracy of the spatial position, and then the inverse depth information is obtained as the initial information of the inverse depth estimation.
b. Using the normalized eight-point algorithm, the basic matrix in epipolar geometry is obtained, and then the pose of the camera is obtained as the initial value of optimization.
c. The pose of monocular camera is modeled by a factor graph model, and the pose estimation is transformed into an unconstrained optimization problem by using the transformation relationship between Lie group and Lie algebra. Then the optimized camera pose is obtained, which is ready for further inverse depth estimation.
d. Using the inverse depth of Gauss-Uniform mixture probability distribution based on probability graph model and approximate inference, the recurrence formula can be obtained, which can facilitate sequential processing of multiple images.
e. The depth information is quantitatively measured and compared by using TUM datasets, and the length of space object is measured by using inverse depth information, thus the measurement accuracy of this method is indirectly verified.

The above steps are iterative. Generally, the more sequential images are available, the higher the estimation accuracy will be.

III. OPTIMIZING CAMERA POSTURE USING FACTOR GRAPH MODEL

Fig. 3 is a schematic diagram of the factor model, in which each node represents a relevant random variable, where x_i

represents the node about position and attitude of the camera, l_i is the landmark node, o_{ij} is the observation node of the jth landmark of the ith camera, f_i is the factor function indicating the relationship between the position and attitude of the ith camera and the 1st camera. In the figure, the positions of all cameras are relative positions to that of the first camera. Suppose that the factor function satisfies the following equation of Gaussian distribution:

$$f_i(x_i) = N\left(h(x_i)\,|\mu_i, \sigma_i^2\right) = \frac{1}{\sqrt{2\pi * \sigma_i^2}} \exp\left(\frac{(h(x_i) - \mu_i)^2}{2\sigma_i^2}\right) \tag{26}$$

where, $h(x_i)$ is the re-projection of the ith camera, μ is the corresponding pixel coordinate, $h(x_i) - \mu_i$ is the re-projection error, σ_i^2 is the variance of re-projection.

Therefore, the joint probability distribution of all variables can be written in the form of a product of factors:

$$F(X) = \prod_i f_i(x_i) \tag{27}$$

The maximum posteriori probability inference takes the form of:

$$X_{\max} = \arg\max_X F(X) = \arg\max_X \prod_i f_i(x_i) \tag{28}$$

After the logarithm of the upper formula is taken, the maximum posteriori probability inference problem can be transformed into the problem of minimizing the sum of nonlinear least squares.

$$X_{\max} = \arg\min_X \frac{1}{2} \sum_i \|h(x_i) - \mu_i\|^2 \tag{29}$$

Supposing that the coordinates of a point in three-dimensional space are $U_i = [X_i, Y_i, Z_i]^T$ and the projection pixel coordinates are $\mu_i = [u_i, v_i]^T$, we can calculate the pose of the camera, R and t, using n space points. Supposing that its Lie algebra is ξ, we can obtain the relationship between the position of the pixel and the position of the space point as follows:

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = KT \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} = K \exp\left(\xi^\wedge\right) \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{30}$$

Written as a matrix:

$$s_i \mu_i = K \exp\left(\xi^\wedge\right) U_i \tag{31}$$

Therefore, the re-projection function is:

$$h(x_i) = \mu_i = \frac{1}{s_i} K \exp\left(\xi^\wedge\right) U_i \tag{32}$$

By linearizing the re-projection function, we can get:

$$h(x_i) = h\left(x_i^0 + \Delta x_i\right) = h\left(x_i^0\right) + J\Delta x_i \tag{33}$$

where, J is Jacobian matrix, $\Delta x_i = x_i - x_i^0$, in which, x_i^0 is initial posture.

Therefore, the error function can be expressed as follows:

$$\begin{cases} e^2 = \left\|h\left(x_i^0\right) + J\Delta x_i - \mu_i\right\|^2 = \|J\Delta x_i - b\|^2 \\ b = \mu_i - h\left(x_i^0\right) \end{cases} \tag{34}$$

Nonlinear minimization problems can be transformed into:

$$X_{\max} = \arg\min_X \frac{1}{2} \sum_i \|J\Delta x_i - b\|^2 \tag{35}$$

For the derivation of the upper formula with respect to Δx_i, the normal equation can be obtained as follows:

$$\left(J^T J\right) \Delta x_i = J^T b \tag{36}$$

In order to improve the efficiency of the algorithm, the Levenberg-Marquart (LM) method is used to modify the normal equation. The LM method allows multiple iterations to converge, and the step size is controlled in the execution region. This method is also known as the trust region method. The revised equation is as follows:

$$\left(J^T J + \lambda I\right) \Delta x_i = J^T b \tag{37}$$

where, I is the unit matrix of order n, and λ is a positive real number. When $\lambda = 0$, the method becomes the Gauss-Newton method; when λ is large, $\Delta x_i \approx \frac{1}{\lambda} J^T b$, it is to update along the direction of the negative gradient of the error function.

The next key problem is to solve Jacobian matrix, J. Because the rotation matrix has orthogonal and determinant constraints of 1, when R and t are used as optimization variables, additional variable constraints will be introduced, which increases the difficulty of optimization problems. Therefore, the pose estimation can be transformed into an unconstrained optimization problem using the transformation relationship between Lie group and Lie algebra, to conveniently solve the optimal problem.

If transformed into the camera coordinate system, the coordinate of the space point is $U' = \left[X', Y', Z'\right]^T$, namely,

$$U' = \exp\left(\xi^\wedge\right) U_i = \left[X', Y', Z'\right]^T \tag{38}$$

Therefore, the camera projection model can be transformed into:

$$s_i \mu_i = KU' = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} \tag{39}$$

By using the third line of the equation to eliminate the proportional coefficient s_i, we can get:

$$\begin{cases} u_i = f_x \dfrac{X'}{Z'} + c_x \\ v_i = f_y \dfrac{Y'}{Z'} + c_y \end{cases} \tag{40}$$

Therefore, the error function is:

$$\begin{cases} e_u = f_x \dfrac{X'}{Z'} + c_x - u_i \\ e_v = f_y \dfrac{Y'}{Z'} + c_y - v_i \end{cases} \tag{41}$$

Using the chain rule to derive the position and posture, we can get the result.

$$\frac{\partial e}{\partial \delta\xi} = \frac{\partial e}{\partial U'} \frac{\partial U'}{\partial \delta\xi} \tag{42}$$

The first term is the derivative of the error with respect to the projection point. The derivative of formula (41) can be obtained as follows:

$$\frac{\partial e}{\partial U'} = \begin{bmatrix} \frac{\partial e_u}{\partial X'} & \frac{\partial e_u}{\partial Y'} & \frac{\partial e_u}{\partial Z'} \\ \frac{\partial e_v}{\partial X'} & \frac{\partial e_v}{\partial Y'} & \frac{\partial e_v}{\partial Z'} \end{bmatrix} = \begin{bmatrix} \frac{f_x}{Z'} & 0 & -\frac{f_x X'}{Z'^2} \\ 0 & \frac{f_y}{Z'} & -\frac{f_y Y'}{Z'^2} \end{bmatrix} \tag{43}$$

The second term is the derivative of the transformed point on Lie algebra, the transformation of space points is $T = \exp\left(\xi^\wedge\right)$, and T is multiplied by a perturbation quantity $\Delta T = \exp\left(\delta\xi^\wedge\right)$. Assuming that the Lie algebra of the perturbation term is $\delta\xi = [\delta\rho, \delta\phi]^T$, the derivation process is as follows:

$$\begin{aligned} \frac{\partial (TU)}{\partial \delta\xi} &= \lim_{\delta\xi \to 0} \frac{\exp\left(\delta\xi^\wedge\right) \exp\left(\xi^\wedge\right) U - \exp\left(\xi^\wedge\right) U}{\delta\xi} \\ &\approx \lim_{\delta\xi \to 0} \frac{\delta\xi^\wedge \exp\left(\xi^\wedge\right) U}{\delta\xi} \\ &= \lim_{\delta\xi \to 0} \frac{\begin{bmatrix} \delta\phi^\wedge & \delta\rho \\ 0 & 0 \end{bmatrix} \begin{bmatrix} RU + t \\ 1 \end{bmatrix}}{\delta\xi} \\ &= \lim_{\delta\xi \to 0} \frac{\begin{bmatrix} \delta\phi^\wedge (RU + t) + \delta\rho \\ 0 \end{bmatrix}}{\delta\xi} \\ &= \begin{bmatrix} I & -(RU + t)^\wedge \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I & U'^\wedge \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{44}$$

The derivative of U' about position and posture can be obtained according to the first three-dimension value groups namely,

$$\frac{\partial U'}{\partial \delta\xi} = \left[I \;\; U'^\wedge \right] = \begin{bmatrix} 1 & 0 & 0 & 0 & -Z' & Y' \\ 0 & 1 & 0 & Z' & 0 & -X' \\ 0 & 0 & 1 & -Y' & X' & 0 \end{bmatrix} \tag{45}$$

After multiplying (43) and (45), the derivative of error with respect to posture can be obtained.

$$\frac{\partial e}{\partial \delta\xi} = \begin{bmatrix} \frac{f_x}{Z'} & 0 & -\frac{f_x X'}{Z'^2} & \frac{f_x X' Y'}{Z'^2} & -f_x - \frac{f_x X'^2}{Z'^2} & \frac{f_x Y'}{Z'} \\ 0 & \frac{f_y}{Z'} & -\frac{f_y Y'}{Z'^2} & f_y + \frac{f_y Y'^2}{Z'^2} & -\frac{f_y X' Y'}{Z'^2} & -\frac{f_y X'}{Z'} \end{bmatrix} \tag{46}$$

Through the above method, the optimized pose of the camera can be obtained, and the preparatory work for the optimization of inverse depth can be done.

IV. ESTIMATION OF INVERSE DEPTH BASED ON GAUSS-UNIFORM MIXED PROBABILITY DISTRIBUTION OF PROBABILITY GRAPH

In inverse depth estimation, the outliers in the data could have a serious impact on the results [32], [33]. The usual practice is to neglect the outliers to avoid adverse effects on the results. If the method of collecting data is unreliable, it is reasonable to neglect the outliers. However, in the case of reliable data acquisition, outliers are important data information which must be considered, otherwise false results will be produced.

A data processing method with good robustness should be insensitive to the model and precise sampling distribution of errors. Even if there are some obvious errors in data, they will not have great impact on the whole conclusion or result. The Gauss-Uniform mixture model is a robust data processing method, in which the Gauss distribution corresponds to the good data sampling part, and the Uniform distribution corresponds to the random interference part. Based on an appropriate allocation proportion, good robustness of the system can be guaranteed.

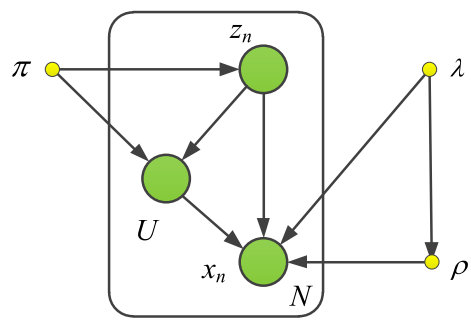

FIGURE 4. Probability diagram of Gauss-Uniform mixed probability distribution model.

Fig. 4 is the probability diagram of the Gauss-Uniform mixture probability distribution model [34]. Assume that $X = \{x_1, \cdots, x_N\}$ is the inverse depth value of sensor observation, $\rho = \{\rho_1, \cdots, \rho_N\}$ is the real inverse depth value, $\pi = \{\pi_1, \cdots, \pi_N\}$ is the proportion of good measurement data, 1-π is the proportion of interference, and the interference signal obeys the Uniform distribution form, $U[\rho_{\min}, \rho_{\max}]$. $\rho_{\min}, \rho_{\max}$ are the minimum inverse depth and the maximum inverse depth measured by the sensor respectively. $\lambda = \{\lambda_1, \cdots, \lambda_N\}$ is the accuracy of Gauss distribution, where $\lambda = \frac{1}{\tau}$, in which τ is the variance of Gauss distribution. Given the true inverse depth ρ, the accuracy λ of the Gauss distribution, and the proportional coefficient π of the correct data, the probability distribution of the inverse depth measured by the mixed model is as follows:

$$p(x_n|\rho_n, \lambda_n, \pi_n) = \pi_n N(x_n|\rho_n, \lambda_n^{-1}) + (1 - \pi_n) U(x_n) \tag{47}$$

All potential discrete variables are recorded as $Z = \{z_{1k}, z_{2k}, \cdots, z_{nk}\}$, in which z_{ik} is a binary random variable. Using the expression of "1-of-K", one element is 1, the rest are 0. Among them, $z_{i1} = 1$, which means the ith measurement value is good measurement data, and $z_{i0} = 0$ which means the ith measurement value is interference data. Therefore, the distribution of potential variables is as follows:

$$\begin{aligned} &p(x_n|\rho_n, \lambda_n, \pi_n, z_{nk}) \\ &= (\pi_n N(x_n|\rho_n, \lambda_n^{-1}))^{z_{nk}}((1-\pi_n)U(x_n))^{1-z_{nk}} \\ &= U(x_n)^{1-z_{nk}} \cdot [N(x_n|\rho_n, \lambda_n^{-1})^{z_{nk}}] \cdot [\pi_n^{z_{nk}}(1-\pi_n)^{1-z_{nk}}] \\ &= C \cdot p(x_n|\rho_n, \lambda_n, z_{nk}) \cdot p(z_{nk}|\pi_n) \end{aligned} \tag{48}$$

where, $C = U(x_n)^{1-z_{nk}}$ is a constant, $p(x_n|\rho_n, \lambda_n, z_{nk}) = N(x_n|\rho_n, \lambda_n^{-1})^{z_{nk}}$ represents conditional probability distribution of observation variables, $p(z_{nk}|\pi_n) = \pi_n^{z_{nk}}(1-\pi_n)^{1-z_{nk}}$ represents conditional probability distribution of mixing coefficients.

We introduce the conjugate prior probability distribution of parameter ρ, λ, π, in which the mixing coefficient π obeys the Beta distribution rule, and its distribution function is as follows:

$$p(\pi_n) = Beta(\pi_n|p_n, q_n) = \frac{\Gamma(p_n+q_n)}{\Gamma(p_n)\Gamma(q_n)}\pi_n^{p_n-1}(1-\pi_n)^{q_n-1} \tag{49}$$

where, Γ is a gamma function, namely, $\mathrm{T}(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt$. According to the important property of gamma functions: $\mathrm{T}(x+1) = x\mathrm{T}(x)$, p_n and q_n can be understood as the reasonable value of experimental data and the number of interference values in the whole experiment process.

The conjugate prior distributions of mean ρ and precision λ are introduced as Gauss-Gamma distribution patterns in the following forms:

$$\begin{cases} p(\rho_n|\lambda_n) = N(\rho_n|\rho_0, (\upsilon_0\lambda_n)^{-1}) \\ p(\lambda_n) = Gam(\lambda_n|a_0, b_0) = \frac{1}{\mathrm{T}(a_0)}b_0^{a_0}\lambda_n^{a_0-1}\exp(-b_0\lambda_n) \end{cases} \tag{50}$$

where, $\rho_0, \upsilon_0, a_0, b_0$ are initial values of Gauss distribution and Gamma distribution respectively.

According to Bayesian theorem, *posterior* $\propto$ *likelihood* $\times$ *prior*, the joint probability distribution of all random variables can be obtained in the form of:

$$\begin{aligned} &P(X, Z, \pi, \rho, \lambda) \\ &\quad = p(X|Z, \rho, \lambda)p(Z|\pi)p(U|Z, \pi)p(\rho|\lambda)p(\lambda)p(\pi) \end{aligned} \tag{51}$$

Assuming a posteriori probability distribution pattern, $p(Z, \rho, \lambda, \pi|X) = q(Z, \rho, \lambda, \pi)$, which is variational, the parameters can be decomposed and estimated by the variational inference method, namely:

$$q(Z, \rho, \lambda, \pi) = q(Z)q(\rho)q(\lambda)q(\pi) \tag{52}$$

Firstly considered is the derivation of the updating equation of factor $q(Z)$. The optimization of probability distribution $q(Z)$ is equivalent to the minimizing of KL divergence, that is, the maximum value of $q(Z)$ appears when KL divergence equals zero. The logarithm of the optimal factor $q^*(Z)$ is:

$$\ln q^*(Z) = \mathrm{E}_{\pi,\rho,\lambda}\left[P(X, Z, \pi, \rho, \lambda)\right] + \text{constant} \tag{53}$$

where, $\mathrm{E}_{\pi,\rho,\lambda}[\cdots]$ is the mathematical expectation of probability distribution of $q(Z)$, which is defined on variable π, ρ, λ.

We are only interested in functions on the right side of the equation, which are related to variable Z. Items unrelated to the variable Z are integrated into the normalized coefficient, and introduce all the expressions into the formulas above to obtain the following formulas:

$$\begin{aligned} &\ln q^*(Z) \\ &= \mathrm{E}_{\rho,\lambda}[\ln p(X|Z, \rho, \lambda)] + \mathrm{E}_{\pi}\left[p(Z|\pi)p(U|Z, \pi)\right] + constant \\ &= \mathrm{E}\left[\sum_{n=1}^{N} z_{nk}\left(-\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\lambda_n - \frac{\lambda_n}{2}(x_n-\rho_n)^2 + \ln\pi_n\right)\right] \\ &\quad + \mathrm{E}\left[\sum_{n=1}^{N}(1-z_{nk})\left(\ln U(x_n) + \ln(1-\pi_n)\right)\right] + constant \\ &= \sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk}\ln\omega_{nk} + constant \end{aligned} \tag{54}$$

$$\omega_{nk} = \begin{cases} \exp\left(\mathrm{E}\left[-\frac{1}{2}\ln 2\pi + \frac{1}{2}\ln\lambda_n - \frac{\lambda_n}{2}(x_n-\rho_n)^2 + \ln\pi_n\right]\right), \\ \qquad z_{nk} = 1 \\ \exp\left(\mathrm{E}\left[\ln U(x_n) + \ln(1-\pi_n)\right]\right), \\ \qquad z_{nk} = 0 \end{cases} \tag{55}$$

The probability distribution is normalized, that is, for all the n values, the sum of all k values is 1. By taking exponents at both ends, we can get:

$$\begin{cases} q^*(Z) = \prod_{n=1}^{N}\prod_{k=1}^{K} r_{nk}^{z_{nk}} \\ r_{nk} = \dfrac{\omega_{nk}}{\sum_{i=1}^{K}\omega_{ni}} \end{cases} \tag{56}$$

We call r_{nk} responsibility which plays an important role in posterior probability distribution. For discrete probability distribution $q^*(Z)$, the following results can be obtained:

$$\mathrm{E}\left[z_{nk}\right] = r_{nk} \tag{57}$$

The statistics defining responsibility for observations are as follows:

$$\begin{cases} N_k = \sum_{n=1}^{N} r_{nk} \\ \bar{x}_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}x_n \\ S_k = \frac{1}{N_k}\sum_{n=1}^{N} r_{nk}x_n^2 - \bar{x}_k^2 \end{cases} \tag{58}$$

The logarithm of the optimal factor $q^*(\pi)$ is:

$$\begin{aligned}
&\ln q^*(\pi) \\
&= \mathrm{E}_Z\left[\ln p(Z|\pi) + \ln p(\pi)\right] + \text{constant} \\
&= \mathrm{E}\left[\sum_{n=1}^{N} z_{nk}\ln\pi_n + (1-z_{nk})\ln(1-\pi_n)\right] \\
&\quad + (p_n-1)\ln\pi_n + (q_n-1)\ln(1-\pi_n) + \text{constant} \\
&= \mathrm{E}\left[\left(p_n + \sum_{n=1}^{N} z_{nk} - 1\right)\ln\pi_n \right. \\
&\quad \left. + \left(q_n + \sum_{n=1}^{N}(1-z_{nk}) - 1\right)\ln(1-\pi_n)\right] + \text{constant}
\end{aligned} \tag{59}$$

Therefore, the updating equation of Beta distribution parameters is:

$$\begin{cases} p_{n+1} = p_n + r_{nk} \\ q_{n+1} = q_n + 1 - r_{nk} \end{cases} \tag{60}$$

The logarithm of the optimal factor $q^*(\rho)$ is:

$$\begin{aligned}
&\ln q^*(\rho) \\
&= \mathrm{E}_{Z,\lambda}\left[\ln p(X|Z,\rho,\lambda)\right] + \mathrm{E}_\lambda\left[\ln p(\rho|\lambda)\right] + constant \\
&= -\frac{\mathrm{E}[\lambda]}{2}\left[\sum_{n=1}^{N} r_{nk}(x_n-\rho)^2 + \upsilon_0(\rho-\rho_0)^2\right] + constant
\end{aligned} \tag{61}$$

With respect to ρ squaring, the Gauss distribution $N(\rho|\rho_N, \lambda_N^{-1})$ can be obtained, where the mean and variance are respectively as follows:

$$\begin{cases} \rho_N = \dfrac{\upsilon_0\rho_0 + N_k\bar{x}_k}{\upsilon_0 + N_k} \\ \lambda_N = (\upsilon_0 + N_k)\,\mathrm{E}[\lambda] \end{cases} \tag{62}$$

The logarithm of the optimal factor $q^*(\lambda)$ is:

$$\begin{aligned}
&\ln q^*(\lambda) \\
&= \mathrm{E}_{Z,\rho}\left[\ln p(X|Z,\rho,\lambda)\right] + \mathrm{E}_\rho\left[\ln p(\rho|\lambda)\right] \\
&\quad + \ln p(\lambda) + constant \\
&= -\frac{\lambda}{2}\mathrm{E}_\rho\left[\sum_{n=1}^{N} r_{nk}(x_n-\rho)^2 + \upsilon_0(\rho-\rho_0)^2\right] - b_0\lambda \\
&\quad + \left(\frac{N_k+1}{2} + a_0 - 1\right)\ln\lambda + constant
\end{aligned} \tag{63}$$

Therefore, $q(\lambda)$ is a Gamma distribution pattern, $Gam(\lambda|a_N, b_N)$, with the following parameters:

$$\begin{cases} a_N = a_0 + \dfrac{N_k+1}{2} \\ b_N = b_0 + \dfrac{1}{2}\mathrm{E}_\rho\left[\sum_{n=1}^{N} r_{nk}(x_n-\rho)^2 + \upsilon_0(\rho-\rho_0)^2\right] \end{cases} \tag{64}$$

The expression of $\mathrm{E}[\lambda]$, $\mathrm{E}[\rho]$, $\mathrm{E}\left[\rho^2\right]$ is obtained by using the priori distribution without information. Assuming $\rho_0 = \upsilon_0 = a_0 = b_0 = 0$ and the standard result $\mathrm{E}[\lambda] = \frac{a_N}{b_N}$ of the mean value of Gamma distribution, we can get:

$$\begin{cases} \mathrm{E}[\rho] = \bar{x}_k, \mathrm{E}\left[\rho^2\right] = \bar{x}_k^2 + \dfrac{S_k}{N_k}, \mathrm{E}[\lambda] = \dfrac{1}{S_k} \\ b_N = b_0 + \dfrac{1}{2}\left[N_kS_k + \upsilon_0\left(\bar{x}_k^2 + \dfrac{S_k}{N_k} - 2\rho_0\bar{x}_k + \rho_0^2\right)\right] \end{cases} \tag{65}$$

V. EXPERIMENT

In order to ensure the comparability of the experiment, all the camera positions and postures were optimized using the g2o method. The depth filter, inverse depth filter and mixed probability distribution inverse depth filter were compared respectively. The inverse depth filter is for the method mentioned in Section II.C of this paper.

The initialization parameters of Beta distribution are $p_0 = 1$, $q_0 = 1$, and the ratio coefficient is $\pi_0 = \frac{p_0}{p_0+q_0} = 0.5$. The initialization parameters of Gamma distribution are $a_0 = 1$, $b_0 = 1$. The camera pose obtained by the classical eight-point method was used as the initial value for factor diagram optimization. The camera depth information obtained using the triangle method and the location of the space obtained using the first order geometric correction method were used as initial values for Gauss-Uniform mixture probability distribution.

A. QUANTITATIVE CONTRAST EXPERIMENT OF DEPTH ESTIMATION

In this paper, we used RGB-D Benchmark in TUM datasets to carry out experiments. The RGB images and their corresponding depth maps are shown in Fig. 5, respectively.

In the experiment, we compared the depth obtained by the mixed inverse depth estimation method based on probability graph with the true value of depth provided by TUM, and obtained the quantitative comparison results. Thirty images were selected from TUM datasets, and then the same 30 feature points were extracted from 30 images. The depth values of corresponding feature points were calculated and compared with the true values. All relative errors were averaged as the overall error of image depth estimation by this method.

In this paper, the depth estimation method, the inverse depth estimation method and the inverse depth estimation method based on mixed probability model were used to carry out experiments. The overall errors of the three methods are shown in TABLE 1, respectively.

From TABLE 1, it can be seen that the accuracy of depth estimation by inverse depth method is better than that by depth method, because the assumption of inverse depth makes the distribution of measurement data more reasonable, and therefore improves the measurement accuracy. The precision of mixed probability inverse depth estimation method is better than that of inverse depth estimation method, because the

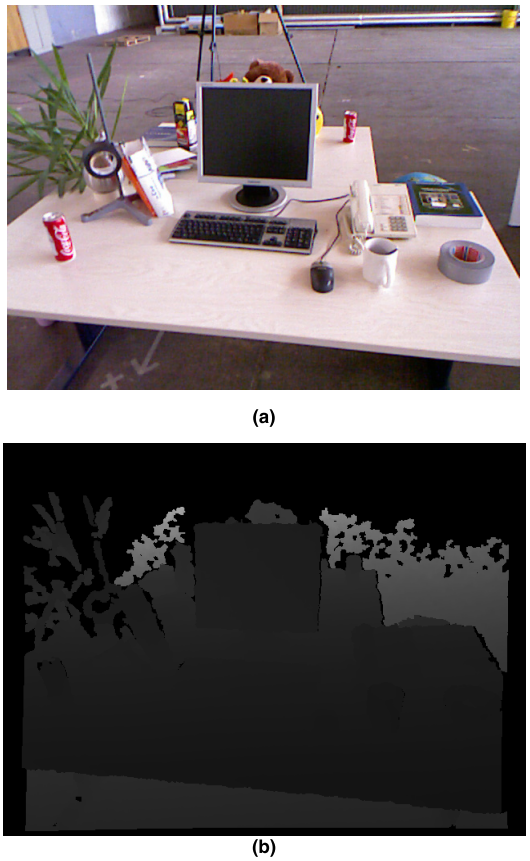

FIGURE 5. RGB images and depth maps used in experiments. (a) RGB images used in experiments. (b) Depth image.

TABLE 1. Overall error contrast results.

Method	Depth Method	Inverse Depth Method	Mixed Probability Inverse Depth Method
Error	16.71%	13.47%	7.61%

mixed probability distribution improves the robustness of the system, and can make more reasonable and full use of all measured data.

In order to verify the applicability of this method to indoor and outdoor environments, 720P USB camera were used to carry out experiments, and the focal length of the camera is 3.6mm.In the experiment, the camera calibration results are as follows: $f_x = 1500.0, f_y = 1510.5$, $u_x = 708.7, u_y = 338.7$. Thirty groups of photographs of indoor and outdoor environment were collected for experiment. Using the depth information of space points, the length of space objects can be calculated, and then the measurement accuracy of the estimation method can be obtained. This is also an application of depth information.

B. ESTIMATION AND CONTRAST TESTING FOR INDOOR ENVIRONMENTS

In Fig. 6, (a) is the original image of indoor environment, in which the really measured distance is 0.4m; (b) is the depth map obtained by depth filter filtering, in which the distance measured by depth map is 0.27m, which is with the largest error; (c) is the inverse depth image filtered by Gauss inverse depth filter, in which the outlines of the objects can be clearly seen and the measured distance is 0.33m, which is with a small error; (d) is a mixed inverse depth filtering

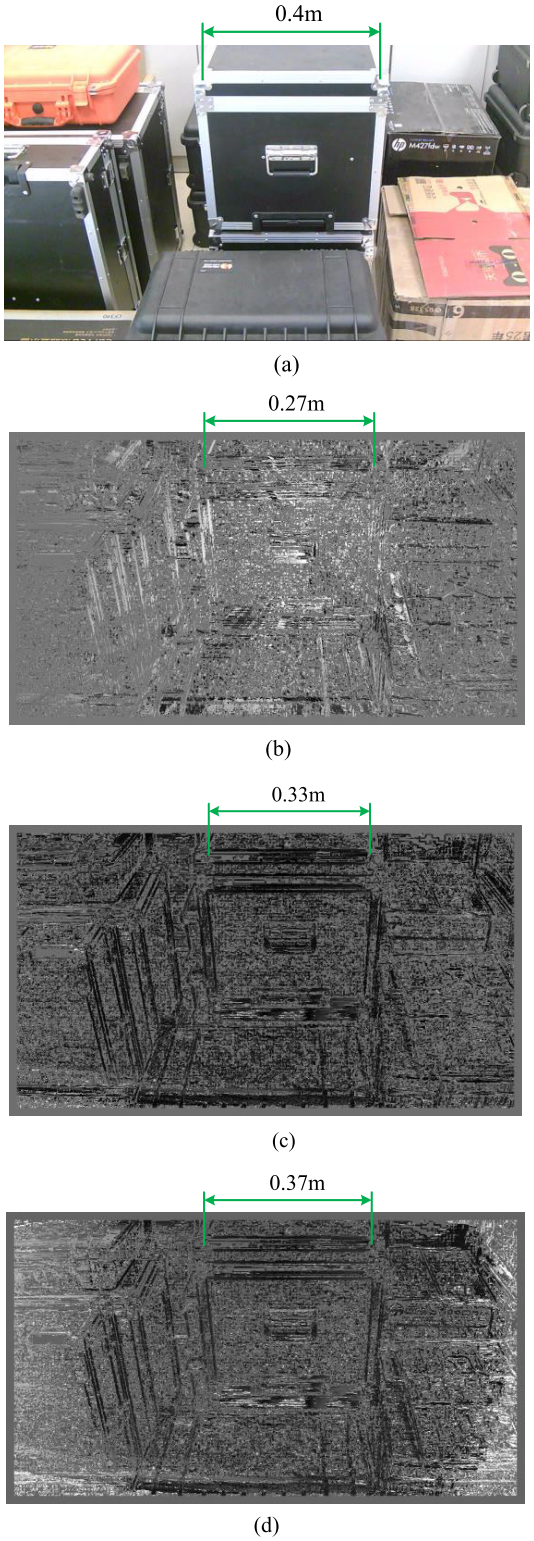

FIGURE 6. Contrast chart for indoor environment experiment. (a) Source image. (b) Depth map error:32.5%. (c) Inverse depth map error:17.5%. (d) Mixed inverse depth map error:7.5%.

inverse depth map that is based on the probability graph, in which the outlines of the object are clearly visible and the measured distance is 0.37m, which is with the minimum error.

By comparing the calculation results of depth filter and the inverse depth filter, it can be concluded that the calculation of inverse depth filter is more stable. For the depth filter, it's assumed that the depth of the pixels near the measurement point agrees with the Gauss distribution, but this assumption will suffer that the pixels close to the center of the camera are too concentrated, and those far from the center of the camera are tailed, thus resulting in uneven data distribution and poor anti-jamming ability. On the contrary, the inverse depth filter take the reciprocal of the depth information of the pixels assuming that its reciprocal satisfies the Gauss distribution. This hypothesis effectively solves the problem of pixel tailing, disperses the pixels near the camera, and makes the distribution of the depth of the pixel more reasonable. Therefore, its experimental result is more stable and anti-interference.

As can be seen from the above figure, the filtering effect of mixed inverse depth filter based on probability graph is better, more stable and more robust than that of the Gauss inverse depth filter. In the mixed model, besides assuming the introduction of Gaussian distribution near the measurement point, random interference is also introduced according to the prior information of the inverse depth. The optimal mixing coefficient can be obtained by approximate inference. Reasonable mixing distribution can not only improve the measurement accuracy, but also enhance the robustness of the system.

C. ESTIMATION AND CONTRAST TESTING FOR OUTDOOR ENVIRONMENTS

In Fig. 7, (a) is the original picture of the outdoor environment, where the really measured distance between the two points is 1m. (b) is the depth map obtained after filtering of the depth filter, in which the measured distance is 0.77m and the error is relatively large. (c) is the inverse depth map obtained after filtering of the inverse depth filter, in which the measuring distance is 0.8m, of which the accuracy is improved and the outlines are obvious, compared with the depth method. (d) is the inverse depth map obtained after filtering of the mixed inverse depth filter, based on the probability graph, in which the measured distance is 1.05m, the error is relatively small, and the outlines are clear.

By comparing the results of the above three experiments, it can be seen that the mixed inverse depth filtering method based on probability graph is superior to the inverse depth filtering method, and the inverse depth filtering method is superior to the depth filtering method. The inverse depth filtering mainly solves the problem of data tailing, while the mixed inverse depth filtering mainly solves the problem of anti-jamming. Generally speaking, the mixed inverse depth filtering method not only improves the robustness of data, but also obtains measurement results with less data tailing.

Usually, indoor environments are relatively stable, the light intensity is stable, and the interference factor is relatively small. Therefore, stable feature extraction and image matching is ensured, and the measurement error is relatively small. An outdoor environment is changeable, and its intensity of illumination fluctuates with the position of measurement,

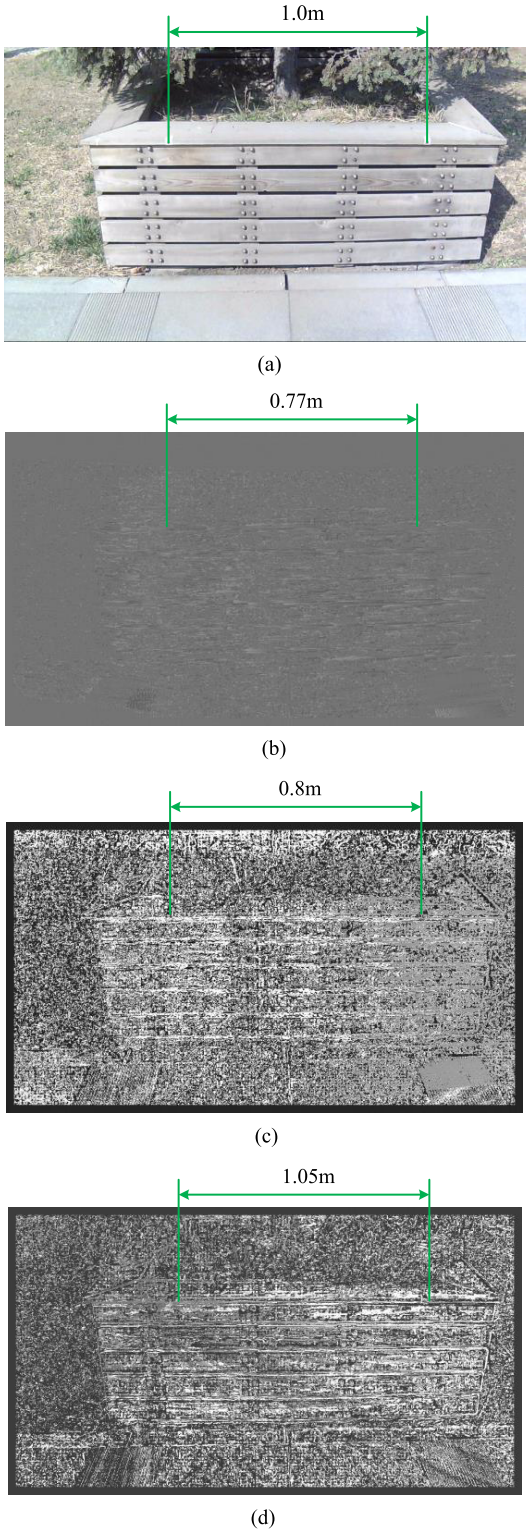

FIGURE 7. Contrast chart for outdoor environment experiment. (a) Source image. (b) Depth map error:23%. (c) Inverse depth map error:20%. (d) Mixed inverse depth map error:5%.

which results in a large difference in gray value of the same feature point in different images, and it has a certain randomness which makes it difficult for feature extraction and image matching. The mixed inverse depth filtering method based on probability graph can deal with the errors caused by random interference and improve the robustness of the measurement. The experimental results show that the mixed inverse depth filtering method has better measurement results in both indoor and outdoor environments.

D. EFFECT OF MIXED PROPORTION COEFFICIENT ON EXPERIMENTAL RESULTS

This experiment mainly adjusted the proportionality coefficient of Gauss-Uniform Mixing Probability Distribution, while other parameters remained unchanged. The variation of the measured distance under different experimental environments was observed. The experimental results are shown in TABLE 2.

TABLE 2. Experimental results of mixed proportional coefficient.

Proportion	Indoor environment			Outdoor environment		
	True value	Measured value	Error	True value	Measured value	Error
25%	0.4m	0.37m	7.5%	1m	1.15m	15%
40%	0.4m	0.37m	7.5%	1m	1.09m	9%
50%	0.4m	0.37m	7.5%	1m	1.05m	5%
60%	0.4m	0.37m	7.5%	1m	1.02m	2%
75%	0.4m	0.37m	7.5%	1m	0.96m	4%

From TABLE 2, we can see that the change of proportion coefficient has little effects on the indoor environment experiment. The main reason is that indoor environment is relatively stable, illumination intensity is constant and random interference is relatively small. That is to say, the proportion of the interference factor in the experiment is quite small. With the increase of the amount of experiments, the proportion of the interference factor decrease, and ultimately, their effects on the experimental results can almost be neglected.

However, the proportion coefficient has a significant effect on outdoor environmental experiments. With the decrease of the proportion of the interference factor, the measurement accuracy first increases, but after reaching 40%, the accuracy becomes worse. The main reason is that the outdoor environment is susceptible to changes of light intensity and random noises. The influence of interference factors on the experimental results can't be ignored. When the proportion of interference is relatively large, it can seriously affect the measurement of real data, so the error may be relatively large. With the reduction of interference, the measurement accuracy will also be improved. When the interference disappears gradually, the measurement accuracy becomes worse. The main reason is that when the interference is reduced to a certain extent, the mixture probability distribution matches the Gauss probability distribution, and the advantage of the mixture distribution will disappear, so the accuracy becomes worse.

The experimental results show that when the interference factor is less than 50%, the measurement accuracy of mixed probability distribution is relatively high.

VI. CONCLUSION

Initialization information has a great impact on experimental results. It is very important to select good initial values for optimization. In this paper, the position and attitude obtained using the normalized eight-point method were taken as the initial values for the factor graph model optimization, and the inverse depth information obtained using the triangle method was taken as the initial value for inverse depth information estimation by Gauss-Uniform Mixed Probability Distribution Model.

The application range of the depth estimation is limited; the data near the center of the camera is zero, and the phenomenon of data tailing occurs at distance far away from the camera, resulting in abnormal Gauss distribution, and an estimation error relatively large. In order to solve above problems, this paper proposed a method of inverse depth distribution based on Gauss, which takes reciprocal of depth information and assumes that the reciprocal obeys Gauss distribution. This method can effectively solve the problem of data tailing and improve the accuracy and stability of depth estimation.

In this paper, a factor graph model was introduced into the inverse depth estimation, making the idea of using a factor graph to represent inverse depth estimation more clear and resulting in more potential applications of the method and better accuracy of the estimation. In the process of pose optimization, in order to cope with the derivation of rotation matrix, additional variable constraints were introduced, which increases the difficulty of optimization. Thus, the transformation relationship between Lie group and Lie algebra can be used to transform the pose estimation into an unconstrained optimization problem, thus facilitating the solution of the optimal problem.

Aiming to solve the problem that the outliers in data could seriously affect the results of inverse depth estimation, this paper proposed an inverse depth estimation method based on the Gauss-Uniform mixed probability distribution model, in which the Gauss distribution corresponds to the good data sampling part and the Uniform distribution corresponds to the random interference part. The system can be guaranteed to have a good robustness by appropriate allocation proportion. By approximate inference, the recurrence formula can be obtained, which can facilitate sequential processing of multiple images. This method is insensitive to the model of data and the distribution of data sampling; even if some errors in the sampled data are large, no great impacts will be imposed on the results and the conclusions.

In conclusion, the inverse depth estimation method of mixed probability distribution, which is based on probability graph, not only has good robustness, but also has high estimation accuracy.

REFERENCES

[1] M. Pizzoli, C. Forster, and D. Scaramuzza, "REMODE: Probabilistic, monocular dense reconstruction in real time," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May /Jun. 2014, pp. 2609–2616.

[2] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2320–2327.

[3] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May /Jun. 2014, pp. 15–22.

[4] C. Forster, M. Pizzoli, and D. Scaramuzza, "Air-ground localization and map augmentation using monocular dense reconstruction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 3971–3978.

[5] N. Ammann and L. G. Mayo, "Undelayed initialization of inverse depth parameterized landmarks in UKF-SLAM with error state formulation," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatron. (AIM)*, Jul. 2018, pp. 918–923.

[6] S. Zhao and Z. Fang, "Direct depth SLAM: Sparse geometric feature enhanced direct depth SLAM system for low-texture environments," *Sensors*, vol. 18, no. 10, p. 3339, 2018.

[7] C. C. Su, L. K. Cormack, and A. C. Bovik, "Bayesian depth estimation from monocular natural images," *J. Vis.*, vol. 17, no. 5, p. 22, 2017.

[8] J. Civera, A. J. Davison, and J. M. M. Montiel, "Inverse depth parametrization for monocular SLAM," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 932–945, Oct. 2008.

[9] J. Civera, A. J. Davison, and J. M. M. Montiel, "Inverse depth to depthconversion for monocular SLAM," in *Proc. Int. Conf. Robot. Autom.*, 2007, pp. 2778–2783.

[10] C. Joly and P. Rives, "Bearing-only SAM using a minimal inverse depth parametrization-application to omnidirectional SLAM," in *Proc. ICINCO*, Funchal, Madeira, Portugal, 2010, pp. 281–288.

[11] E. Perdices and J. M. Cañas, "SDVL: Efficient and accurate semi-direct visual localization," *Sensors*, vol. 19, no. 2, p. 302, 2019.

[12] T. Drummond, "Lie groups, Lie algebras, projective geometry and optimization for 3D geometry." *Eng. Comput. Vis.*, 2014. Accessed: Jul. 24, 2018. [Online]. Available: https://twd20g.blogspot.com/2014/02/updated-notes-on-lie-groups.html

[13] Y. Ma *et al.*, *An Invitation to 3-D Vision: From Images to Geometric Models*, vol. 26. Springer, 2012.

[14] G. S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups: Analytic Methods and Modern Applications*, vol. 2. Springer, 2011.

[15] D. Lee and H. Myung, "Solution to the SLAM problem in low dynamic environments using a pose graph and an RGB-D sensor," *Sensors*, vol. 14, no. 7, pp. 12467–12496, 2014.

[16] G. Dubbelman and B. Browning, "COP-SLAM: Closed-form online pose-chain optimization for visual SLAM," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1194–1213, Oct. 2015.

[17] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Inst. Technol., Atlanta, GA, USA, Tech. Rep. GT-RIM-CP&R-2012-002, 2012.

[18] F. Dellaert and M. Kaess, "Square Root SAM: Simultaneous localization and mapping via square root information smoothing," *Int. J. Robot. Res.*, vol. 25, no. 12, pp. 1181–1203, 2006.

[19] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Int. J. Robot. Res.*, vol. 31, no. 2, pp. 216–235, 2012.

[20] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. Robot.*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[21] M. Kaess, A. Ranganathan, and F. Dellaert, "Fast incremental square root information smoothing," in *Proc. IJCAI*, Jan. 2007, pp. 2129–2134.

[22] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous Robot Vehicles*. New York, NY, USA: Springer, 1990, pp. 167–193.

[23] G. Vogiatzis and C. Hernández, "Video-based, real-time multi-view stereo," *Image Vis. Comput.*, vol. 29, no. 7, pp. 434–441, 2011.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[25] P. Coretto and C. Hennig, "Maximum likelihood estimation of heterogeneous mixtures of Gaussian and uniform distributions," *J. Stat. Planning Inference*, vol. 141, no. 1, pp. 462–473, 2011.

[26] R. P. Browne, P. D. McNicholas, and M. D. Sparling, "Model-based learning using a mixture of mixtures of Gaussian and uniform distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 814–817, Apr. 2012.

[27] G. Bottegal, A. Y. Aravkin, H. Hjalmarsson, and G. Pillonetto, "Outlier robust system identification: A Bayesian kernel-based approach," *IFAC Proc.*, vol. 47, no. 3, pp. 1073–1078, 2014.

[28] A. Aguilar-González, M. Arias-Estrada, and F. Berry, "Depth from a motion algorithm and a hardware architecture for smart cameras," *Sensors*, vol. 19, no. 1, p. 53, 2019.

[29] X. Yang, H. Luo, Y. Wu, Y. Gao, C. Liao, and K.-T. Cheng, "Reactive obstacle avoidance of monocular quadrotors with online adapted depth prediction network," *Neurocomputing*, vol. 325, pp. 142–158, Jan. 2019.

[30] H. Ikeoka and T. Hamamoto, "Accuracy improvement of depth estimation with tilted optics by optimizing neural network," in *Proc. Int. Workshop Adv. Image Technol. (IWAIT)*, vol. 11049, Mar. 2019, Art. no. 1104934.

[31] W. Liu, S. Wu, X. Wu, and H. Zhao, "Calibration method based on the image of the absolute quadratic curve," *IEEE Access*, vol. 7, pp. 29856–29868, 2019.

[32] Y. Wang, W. Fu, and D. P. Agrawal, "Gaussian versus uniform distribution for intrusion detection in wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 2, pp. 342–355, Feb. 2013.

[33] K. Tanaka and A. Takemura, "Strong consistency of MLE for finite mixtures of location-scale distributions when the scale parameters are exponentially small," May 2006, *arXiv:math/0605148*. [Online]. Available: https://arxiv.org/abs/math/0605148

[34] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

WENLEI LIU was born in Yantai, China. He received the B.S. degree from the Shandong University of Science and Technology, Qingdao, in 2012, and the M.S. degree from Beihang University, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree.

He mainly studies high precision visual relative navigation and the data fusion of multi data sensors to achieve the purpose of cooperative detection and guidance.

SENTANG WU received the Ph.D. degree in dynamics, ballistics, and aircraft motion control systems from National Aviation University, Ukraine, in 1992.

He is currently a Professor of automation science and electrical engineering and a Ph.D. Tutor with Beihang University, Beijing, China. He is also the Navy Missile Expert with the National Defense Basic Research Institute and a member of the academic committee. His research interests include the theory and the application of nonlinear stochastic systems, computer information processing and control, aircraft cooperative control, precision, and guidance.

XIAOLONG WU was born in Chengdu, China, in 1988. He received the B.S. degree from Sichuan University, Chengdu, in 2010, and the Ph.D. degree from Beihang University, Beijing, in 2017.

He is with the Navigation and Control Technology Institute of NORINCO Group. He is currently a Research Associate. His research interests include overall design of the guidance and control system for unmanned aerial vehicles, autonomous decision and online planning, and application of computer vision.

KAI LI received the bachelor's degree from the Nanjing University of Aeronautics and Astronautics, China, in 2016, and the master's degree in aerospace engineering from the Harbin Institute of Technology, China, in 2018. He is currently pursuing the Ph.D. degree in control science and engineering from Beihang University, China. His research interests include multi-agent networks and aircraft cooperative control.