Abstract—Autonomous aerial refueling (AAR) is a crucial technique of unmanned aerial vehicles (UAVs) to push the fuel limits and play a great role in both civilian and military domains. This paper presents an accurate and robust binocular pose estimation algorithms optimized by brain storm optimization (BSO), which is developed from a robust non-iterative solution of PnP (RPnP). In this algorithm, BSO is employed to select the best rotation axis in RPnP. A large quantity of contrastive simulation experiments has been conducted to verify the proposed algorithm. Furthermore, this work built an aerial verification platform for vision-based AAR. A tanker UAV and a receiver UAV were applied to implement AAR. The real-time visual measuring system includes feature extraction and pose estimation. Several state-of-the-art pose estimation algorithms and the proposed method (which refers to BSO-BPnP) have been tested in the aerial verification platform. Adequate comparative trials and detailed analyses are given in this paper.

Keywords—autonomous aerial refueling (AAR), unmanned aerial vehicles (UAVs), pose estimation, brain storm optimization (BSO), binocular camera system

I. INTRODUCTION

Unmanned Aerial Vehicle (UAV) is a type of aircraft that does not require human pilot on board [1]. UAV shows immense potential in both civilian and military domains. Benefiting from its higher cost performance, light size, and maneuvering capabilities, UAV gradually replaces manned aircraft to execute various ordinary and sophisticated missions, such as border patrol [2], surveillance [3], forest firefighting [4, 5], and target detection and traction [6, 7], etc. The limited fuel has been a critical bottleneck for further development[8]. Autonomous aerial refueling (AAR) is a powerful way to solve this problem [9]. Common aerial refueling can be clasped into boom and probe-and-drogue [10]. However, a human operator is quite a challenging task.

A key technology of AAR is gaining the accurate pose information between receiver and tanker. Currently, inertial navigation system (INS) and global positioning system (GPS) are widely used for aircraft navigation[11]. However, INS is not suitable because the accumulation of errors. GPS-based measurements are ill-suited for docking maneuver in autonomous aerial refueling accurately. 3D flash lidar sensor is utilized to measure the relative position and attitude between aircraft for drogue tracking [12]. Boom approach using visual snake optical sensor was proposed in [13]. A real-time visual measurement in simulation environment is designed for semi-autonomous docking within aerial refueling [14]. Computer vision-based navigation could refrain from error accumulation and meet precision requirement. Therefore, visual measuring system is applied for pose estimation for AAR. BPnP is proposed based on the RPnP (a robust non-iterative solution of PnP) [22]. BSO is an innovative Swarm Intelligence (SI) algorithm proposed by Shi in 2009 [16], which is motivated by the brainstorming of human beings. Recently years, BSO has been proven feasible and effective in dealing with various optimization issues [17-19]. In this paper, BSO is implemented to find the optimal rotation axis, and then yield the pose of the binocular camera system.

This paper is organized as follows: Section II introduces the feature extraction of the receptacle. Section III presents the BPnP pose estimation method. BSO- BPnP is given in Section IV. Section V introduces the experimental platform configuration. The simulation results and outfield aerial tests are shown in Section VI.

II. FEATURE EXTRACTION OF THE RECEPTACLE

Feature extraction is a significant step in visual measurement. Firstly, the labeled markers are recognized from image sequences, and the pixel coordinates needs to be
resolved. Then, the pose estimation algorithm is applied to figure out the relation (rotational and translational) between camera coordinate system and receptacle coordinate system.

A. Marker Detection

Feature of color is a considerable feature to implement image segmentation. The Hue-Saturation-Intensity (HSV) space can weaken the influences of illumination intensity[14]. Therefore, Red-Green-Blue (RGB) images are mapped to HSV space. The conversion formulas are shown

\[
h = \begin{cases} 
0 & \text{max} = \min \\
60 \times \frac{g - b}{\max - \min} + 0^\circ & \text{max} = r, g \geq b \\
60 \times \frac{g - b}{\max - \min} + 360^\circ & \text{max} = r, g < b \\
60 \times \frac{b - r}{\max - \min} + 120^\circ & \text{max} = g \\
60 \times \frac{r - g}{\max - \min} + 240^\circ & \text{max} = b
\end{cases}
\]

\[
s = \begin{cases} 
0 & \text{max} = 0 \\
\frac{\max - \min}{\max} & \text{max} \neq 0
\end{cases}
\]

\[
v = \max
\]

where \( h, s, v \) are the hue channel, saturation channel and intensity channel, the red channel, green channel and blue channel are indicated by \( r, g, b \), \( \max \) and \( \min \) are the maximum and minimum of color space respectively. \( h \) and \( s \) are selected for the threshold segmentation. The obtained binary image with noisy point will execute morphological operation. Every connected area of adequate size counts as a feature point. The focuses of feature points are calculated as the image coordinate.

B. Point Matching

Point matching is an important procedure for pose estimation. The mapping relation between image coordinates and world coordinates of actual markers around the receptacle in the receiver UAV need be determined. Assume that detected feature point set is \( P = \{ p_1, p_2, \ldots, p_n \} \). The measured corresponding image coordinate of marker \( j \) is \( p_j = (u_j, v_j) \).

The projective point set \( \hat{P} = \{ \hat{p}_1, \hat{p}_2, \ldots, \hat{p}_n \} \) is obtained from the priori pose information. \( \hat{p}_j = (u_j, v_j) \) is the calculated image coordinate according perspective projection model.

The similarity degree of two points set is defined as a cost function. All the Euclidean distances between two points set reaches the minimum when the two points set matches correctly. A matrix \( \text{Err} \) is shown as

\[
\text{Err} = \begin{bmatrix} 
\hat{d}(\hat{p}_1, p_1) & \hat{d}(\hat{p}_1, p_2) & \cdots & \hat{d}(\hat{p}_1, p_n) \\
\hat{d}(\hat{p}_2, p_1) & \hat{d}(\hat{p}_2, p_2) & \cdots & \hat{d}(\hat{p}_2, p_n) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{d}(\hat{p}_n, p_1) & \hat{d}(\hat{p}_n, p_2) & \cdots & \hat{d}(\hat{p}_n, p_n)
\end{bmatrix}
\]

III. THE BPnP ALGORITHM

The pose between camera coordinate and world coordinate from \( n \) correspondences is a “Perspective-n-Point problem” (PnP) [21]. BPnP originates from RnP, which was proposed with the lowest computational complexity \( \Theta(n) \) by Li et.al [22]. In this section, BPnP will be introduced.

A. Problem Description

The rotation and translation from the left camera to the right camera are denoted by \( [R^r \ t^r]_{3 \times 4} \). \( P_i (i = 1, \ldots, n, n \geq 3) \) are points in a 3D reference frame. Their corresponding projections on normalized left image plane \( p_i (i = 1, \ldots, n) \) and right image plane \( p_i' (i = 1, \ldots, n) \) are shown in Fig. 1.

![Fig. 1. The 3D reference points and their projections on the binocular camera image plane.](image)

The task is to estimate the pose information between two cameras: \( [R^l \ t^l]_{3 \times 4} \) and \( [R^r \ t^r]_{3 \times 4} \). If the transformation relation among two camera frames and the 6DOF of left camera are known, the right camera frame can be got based on the principle of stereo vision as follows

\[
\begin{bmatrix} 
R^l & t^l \nonumber \\
0^T & 1
\end{bmatrix} = \begin{bmatrix} 
R^l & 0^T \nonumber \\
0^T & 1
\end{bmatrix} \begin{bmatrix} 
R^r & t^r \nonumber \\
0^T & 1
\end{bmatrix}
\]

B. Selection of a Rotation Axis

BPnP begins with an edge \( \overline{PP}_i \) selected from all the edges \( \{ \overline{PP}_i \mid i > j, i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\} \} \) in the \( n \)-point set.
as a rotation axis [15]. The edge with longest projection length \( \|p_p\| \) is less affected by noise. Then, BSO is employed to select the best rotation axis.

A new coordinate frame \( O_aX_aY_aZ_a \) related to selected edge \( P_{0a}P_{ja} \) is established. The original point \( O_0 \) is in the middle of \( P_{0a}P_{ja} \), and \( Z_a \) axis points to \( P_{ja}P_{0a} \). Then, the \( n \) reference points are divided into \((n-2)\) subsets, which contain three points \( \{P_{0a}P_{ja}P_{k}\} \) \( \{k \neq 0,k \neq j\} \). According to 3 point constraint, \((n-2)\) polynomials are got as follows

\[
\begin{align*}
    f_1(x) &= \alpha_1 x^4 + \beta_1 x^3 + \chi_1 x^2 + \delta_1 x + \epsilon_1 = 0, \\
    f_2(x) &= \alpha_2 x^4 + \beta_2 x^3 + \chi_2 x^2 + \delta_2 x + \epsilon_2 = 0, \\
    \vdots \\
    f_{n-2}(x) &= \alpha_{n-2} x^4 + \beta_{n-2} x^3 + \chi_{n-2} x^2 + \delta_{n-2} x + \epsilon_{n-2} = 0,
\end{align*}
\]

(6)

After finding the minima of the square sum of the polynomials, the rotation axis is determined by

\[
    Z_a = \frac{P_{0a}P_{ja}}{\|P_{0a}P_{ja}\|}
\]

(7)

C. Derivation of Rotation and Translation

When the \( Z_a \) axis is determined, the rotation matrix from \( O_aX_aY_aZ_a \) to the camera \( O_XY_Z \) can be denoted as

\[
    R' = \text{Rot}(Z_a, \alpha) = \begin{bmatrix}
    r_1 & r_2 & r_3 \\
    r_2 & r_3 & r_4 \\
    r_3 & r_4 & r_5
    \end{bmatrix}
\]

\[
    \begin{bmatrix}
    \cos \alpha & -\sin \alpha & 0 \\
    \sin \alpha & \cos \alpha & 0 \\
    0 & 0 & 1
    \end{bmatrix}
\]

(8)

where \( R \) is an orthogonal rotation matrix, and it has the same elements as \( Z_a \) in its third column, \( \text{rot}(Z_a, \alpha) \) denotes a rotation of \( \alpha \) degree around \( Z_a \) axis. The transformation relation between the 2D projections of left and right image plane is obtained by

\[
\begin{align*}
    z'_p p'_i &= R' P_i + t' \\
    z'_p p'_i' &= R'^T P_i' + R'^T t + t'
\end{align*}
\]

(9)

where \( z'_p \) and \( z'_p' \) are the projection depth of \( i \)th reference point whose 3D coordinates are \( P_i = [X_i, Y_i, Z_i]^T \) in \( O_aX_aY_aZ_a \), \( p'_i = (u_i, v_i, 1)^T \) and \( p'_i = (u'_i, v'_i, 1)^T \) are the normalized coordinates in the left and right image plane, respectively.

Rewrite the (9) to a \( 4n \times 6 \) homogenous linear equation system shown by (10).

\[
\begin{bmatrix}
    A_{4n \times 1} \\
    B_{4n \times 1} \\
    C_{4n \times 1}
\end{bmatrix}
\begin{bmatrix}
    \cos \alpha \\
    \sin \alpha \\
    t'_1 \\
    t'_2 \\
    \end{bmatrix}
= \begin{bmatrix}
    -1 \\
    0 \\
    0 \\
    0 \\
    \end{bmatrix}
\]

(10)

where \( [\cos \alpha \sin \alpha t'_1 t'_2 t'_3 t'_4] \) is obtained by Singular Value Decomposition (SVD) [23], \( \text{C}_{4n \times 1} = [C_1 C_2 C_3 C_4] \),

IV. BSO Optimized BPNP

BSO algorithm is used to solve the rotation axis selection problem of BPNP. In this section, the detailed introduction and the structure diagram of BSO is presented.
The position of the $i$th individual is denoted by $X_i = [x_{i1}, x_{i2}, ..., x_{im}]$, where $m = 2$ is the dimension of each individual. $X'_i = [x'_{i1}, x'_{i2}]$ represents the edge $P_{x_{i1}}P_{x_{i2}}$, which is an essential variable. Firstly, $n$ individuals are randomly generated within the solution space. Then each individual is evaluated by the fitness function which is the total error of BnPnP. The optimization objective is obtaining the minimal total error of BnPnP based on BSO.

A random selected cluster center is replaced by randomly generated individual with the possibility of $p_{oa}$. BSO selects one cluster of two with the possibility of $p_{ob}$ randomly. Then select the center or other individuals of the chosen cluster(s) with the possibility of $p_{ob3}$. After that, the selected individual(s) can be updated based on (11) and (12).

$$x_{old} = \begin{cases} x_i, & \text{one cluster} \\ w_1x_{i1} + w_2x_{i2}, & \text{two cluster} \end{cases}$$ (11)

$$x_{new} = x_{old} + \xi N(\mu, \sigma)$$ (12)

where $w_1$ and $w_2$ are the weight of two clusters. $N(\mu, \sigma)$ is the Gaussian random value with mean $\mu$ and variance $\sigma$. $\xi$ denotes the adjusting factor decreasing as the iteration goes.

$$\xi = r \log \text{sig} \left( \frac{N_{c_{\text{max}}} - N_c}{K} \right)$$ (13)

where $r$ is the random number within $[0, 1]$, $N_{c_{\text{max}}}$ denotes the maximum number of iteration, $N_c$ signifies the current iteration number and $K$ adjusts the slope of the $\text{logsig}$ function. After the new individual is generated, the crossover operation is conducted and the best individual is updated.

The process above repeats until all individuals are created and the iteration proceeds until the maximum number of iteration is met. Finally, the best individual of the algorithm is the optimal edge.

V. USING THE TEMPLATE

To validate the validity of the proposed method, the experimental platform configuration is prepared in this section.

Specifically, two hexcicopters are regarded as the receiver UAV and the tanker UAV in Fig.2. A visual measuring system is installed in the tanker. Real-time vision algorithms are developed and running on the on-board vision computer. The video sequence captured onboard is transmitted to a ground station by a digital image transmission. The data results are also displayed in the ground station. Therefore, we can implement real-time monitor from the view of the tanker. If an exception occurs, operator will take over the control power of UAVs. The data transmission is made up of two XBee modules based on UDP protocol.

The schematic diagram of the aerial verification platform and the connection of each subsystem is presented in Fig. 3.

VI. EXPERIMENTAL RESULTS

In this section, two sets of experiments are designed. One is conducting simulation experiment to compare the robustness and accuracy of BSO-BnPnP against that of some advanced pose estimation algorithms. The other type of experiment is executed to verify the designed visual measuring system for UAV autonomous aerial refueling.

A. Simulation Experiment of BSO-BnPnP

The comparison algorithms are shown.

- MLHLM [26]. One of the best pose estimation methods improved from LHM by minimize the object-space collinearity errors of two cameras. MLHLM can obtain the pose and position of a binocular camera system.
- EPnP. An efficient $O(n)$ noniterative solution of PnP by Lepetit [27]. The core concept is to express the n reference points as a weighted sum of four virtual control points.
• DLT [28]. A direct linear transformation method.
• HOMO. A homography method for planar targets [29].
• RPnP. A robust non-iterative PnP solution by Li [22].
• BPnP. A pose estimation method for binocular camera systems proposed by Gan et al [15], which can be used to estimate the poses of both cameras at the same time.
• BSO-BPnP. BSO algorithm is applied to select the best optimal rotation axis in BPnP. BSO-BPnP is suited to the pose estimation of a binocular camera system.

The experimental configuration is two virtual perspective cameras with 640×480 pixels size, 800 pixels focal length. The relation matrix between two cameras are given by

\[
[R^{\odot} \ t^{\odot}] = \begin{bmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
\end{bmatrix}
\] (13)

To prove the effectiveness of BSO-BPnP, three cases of the 3D points are generated randomly [17]. The reference points are in the range \([-2,2] \times [-2,2] \times [4,8], [-2,2] \times [-2,2] \times 0\] and \([1,2] \times [1,2] \times [4,8]\) randomly. Besides, Gaussian noise with \(\sigma = 3\) is added to their projections on the two camera image planes. The parameters and their detail descriptions of BSO are expressed in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m)</td>
<td>Number of ideas</td>
<td>30</td>
</tr>
<tr>
<td>(N_{\text{max}})</td>
<td>Maximum times of iteration</td>
<td>30</td>
</tr>
<tr>
<td>(K)</td>
<td>Number of clusters</td>
<td>3</td>
</tr>
<tr>
<td>(P_{5a})</td>
<td>Probability to directly update a cluster center</td>
<td>0.2</td>
</tr>
<tr>
<td>(P_{6a})</td>
<td>Probability to choose one cluster</td>
<td>0.8</td>
</tr>
<tr>
<td>(P_{6b})</td>
<td>Probability to select the center of the selected cluster</td>
<td>0.4</td>
</tr>
<tr>
<td>(P_{6c})</td>
<td>Probability to select the centers of the two selected clusters</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Each experiment is conducted for 100 times. The results of the three cases are recorded to make statistics, which are given from Fig.4 to Fig.6.
To coplanar PnP problem, Fig.5 presents the simulation results. LHM, MLHM and EPnP are not suitable for this scenario, while BSO-BPnP, BPnP and RPnP remain stable and highly accurate with the error mainly below five points. BSO-BPnP reaches the best results in the test obviously. As shown in Fig.6, EPnP, LHM and MLHM have bad performance in the quasi-singular case because of local minima problem. Besides, the error of HOMO was too large to make a curve in figure. BSO-BPnP obtains the most stable and accurate result in all of methods.

### B. Aerial Practical Experiment

To verify the feasibility and effectiveness of the visual measuring system, outdoor aerial practical experiments are conducted. Two UAVs are placed at random positions. They take off automatically. The receiver navigated by GPS/INS unit flies to the scheduled position relative to the tanker. When the receiver comes into the view of the binocular camera system, the navigating mode is switched to the visual navigating model.

Fig. 7 shows the experimental results of feature extraction. Seven red markers are detected successfully based on the color feature. The central coordinates of these markers are calculated. Fig. 8 gives the results of monocular vision algorithms with the same image. Fig. 9 illustrates the results of binocular vision methods for two corresponding images. The green circles are projected on the original images based on the measured pose. All projective circles coincide with the red markers in the image sequences. It is verify that the tested monocular and binocular vision algorithms are effective and feasible.
center is close to the original markers if the measurements are exact. Fig. 11(d) gives the re-projection error of monocular vision algorithms. RPnP shows the highest accuracy. LHM is the second. EPnP is most unstable. RPNP is the appropriate monocular vision algorithm for pose estimation among them.

Fig. 10. Experimental results of relative distance estimation based on monocular vision in three orthogonal direction between two UAVs during the flight.

Fig. 11. Experimental results of relative distance estimation based on monocular vision in three orthogonal direction between two UAVs during the flight.

Fig. 12. Experimental results of relative distance estimation based on monocular vision in three orthogonal direction between two UAVs during the flight.

Fig. 13. Experimental results of relative distance estimation based on monocular vision in three orthogonal direction between two UAVs during the flight.
The binocular vision algorithms BSO-BPnP, MLHM and absolute orientation are compared in Fig. 12 -Fig. 13. As shown in Fig.12, there are no significant differences in relative distance estimation among the three methods. However, BSO-BPnP is more stable than others in Fig. 13. Absolute orientation reaches outliers sometime. The comparison of the re-projective error of monocular vision algorithms is given in Fig.13(d). BSO-BPnP and MLHM have the re-projection error less than 2 pixels, which means these two methods could be available binocular algorithms in UAVs AAR.

VII. CONCLUSIONS

An accurate and robust binocular pose estimation algorithms optimized by BSO-BPnP is proposed. It is applied in vision-based UAVs autonomous aerial refueling. To verify the feasibility and effectiveness of BSO-BPnP, an aerial vision-based AAR platform established. Considerable simulation tests and outdoor flight tests are conducted. The proposed BSO-BPnP has shown great robustness and accuracy compared with some state-of-the-art pose estimation algorithms. BSO-BPnP is outclassed from other methods in three test cases in simulation environment. Besides, the flight tests indicates that the verification platform could detect the receptacle and implement relative distance and pose estimation with acceptable accuracy. LHM, MLHM, RPnP and the proposed BSO-BPnP show best performance so that they could be the alternative vision algorithms for UAVs AAR. In the future, we will focus on the improvement of the visual measurement in extreme conditions, such as target occlusion, strong exposure and lighting environment. Furthermore, we will improve the aerial verification platform and implement the docking phase successfully and steadily.

REFERENCES


