Double-Robot Dynamic Stereo Vision Algorithm based on UKF and its Application in Moving Target Tracking

Huiying Dong, Xue Lu, Yuqing He, Shengfu Chen, Jianda Han

Abstract

Cooperation between air and ground robots has gradually been one of the latest research directions in robotics, wherein position of ground robot assisted by air robots is a key problem to be solved. This paper, therefore, mainly focuses on the problem of high precise 3D moving target observation. Firstly, a new so-called dynamical stereo-vision system, composed of multiple air robots with vision sensor mounted, is proposed and the corresponding observation algorithm is studied. Secondly, in order to improve the observation precision and realize persistent observation, UKF algorithm is utilized to form a new method for moving target observation and tracking. Subsequently, formation control algorithm is designed to make the air robots keep tracking the interested moving target. Finally, experiments are conducted with respect to a new-designed platform, called indoor multiple-rotor-craft-robot system, to verify the feasibility and validity of the study.

Keywords: State Estimation, Stereo Vision, Tracking, UKF

1. Introduction

With the development of society, people propose the increasingly high requirements about the autonomy of robots (especially mobile robots), and the robot should have the ability to autonomously complete various missions in complex environments. For unmanned aerial robots and ground mobile robots have more complementary features, they are more suitable for collaboration to complete its mission in complex outdoor environment. In recent years, scholars have gradually raised the Air-Ground Robotics cooperation for disaster relief [1], post-disaster information collection [2], clearance [3] and many other applications. It requires ground robots can make high-precision ground-dimensional dynamic positioning with the help of aerial robots [16, 17].

Currently, the available three-dimensional observation methods mostly include ultrasonic, radar, infrared, visual, etc. [4-6] Comparing to other types of measured method, we can find visual observation has many advantages. It has low noise, large dynamic range and accurate measurement of light. Its algorithm is a large flexible and adaptable. In addition, vision itself has information-rich features, which can obtain important signal that other methods can’t perceive. Therefore, the observation based on visual has become one of the most common methods. Vision-based target measurement can be divided into monocular vision and stereo vision [7]. Compared with monocular vision, stereo vision improves the accuracy and robustness of the algorithm by data fusion. In addition, multi-purpose stereo vision also provides a certain amount of redundancy to improve reliability of measurement. When the visual sensor does not detect a target or fails, the sensors are still able to provide three-dimensional information to maintain the observed performance of the entire system [8] [9].

As for the dynamic estimation problem, common methods are Kalman filter (KF), extended Kalman filter (EKF), particle filter (PF), colorless Kalman filter (UKF) and so on. Julier etc. [10] [11] proposed a method, which is based on Unscented transform UKF. The method has the same computational complexity O (L3) of EKF. It has better real-time and does not require linearization. Only it approximates the probability distribution of the nonlinear function using limited sample points.

Some scholars began to study observing moving target by using multi-robot. Such as, Pedro Pinheiro etc. [12, 13] improved the robustness of the system by eliminating inaccurate observation information of some robots.

Taking it as the starting point, this article proposes a so-called dynamic stereo vision system to achieve the three-dimensional observation of the target in dynamic environment. The system is formed...
by two robots equipped with monocular vision sensor. Also, because the process of observation (observation equation) has a strong nonlinearity, this paper proposes a method, that is to achieve three-dimensional motion estimation of moving targets by using UKF algorithm. Finally, in order to validate the system and the feasibility of practical application of the algorithm, we have developed three-dimensional experimental platform of multi-rotor flying robot to observation indoor, and made full experiment to validate the algorithm.

2. Problems statement

In this paper, the study background is two air-robot cooperative observing and tracking a dynamic target. Figure 1 is its schematic. In Figure 1, Mobile Robot is moving target to be observed, each aerial robot installs the CCD. The CCD 1 and CCD 2 are formed a dynamic three-dimensional vision system to observe and track the target. The configuration of each aerial robot’s camera is shown in Figure 2.

Where, \( O_{X_aY_aZ_a} \) is the body coordinate system (camera coordinate system). \( O_{X_wY_wZ_w} \) is the world coordinate system. XY is image coordinate system. P is the target to be observed. \( P' \) is the target for the projection of the target in the image. In order to achieve dynamic tracking and positioning on the ground coordinate system, there needs to address the following four questions. 1) quickly identify the target location in the vision sensor; 2) transform ground targets observed by vision sensor to the world coordinate system, then obtain the absolute position of the target; 3) achieve information observed fusion and dynamic tracked by using the estimation method; 4) control the movement of the robot system to ensure that the target is always located within the field of the visual observation system. This article will address the latter three questions.

3. Dynamic stereo vision algorithm

First, we give the transformation matrixes between the different coordinate systems. The relationship between the image coordinate system and camera coordinate system (body coordinate system) are:

\[
\begin{bmatrix}
    u_i \\
    v_i \\
    1
\end{bmatrix} =
\begin{bmatrix}
    f_{xi} & 0 & u_{0i} & 0 \\
    0 & f_{yi} & v_{0i} & 0 \\
    0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
    X_{ci} \\
    Y_{ci} \\
    Z_{ci} \\
    1
\end{bmatrix}
\] (1)

Where, \( u_i \) and \( v_i \) are the target in the image coordinates of the CCD-i; \( X_{ci}, Y_{ci}, Z_{ci} \) are three-dimensional coordinates of the target in CCD-i coordinate system. \( f_{xi}, f_{yi} \) are the internal reference of cameras (respectively show X and Y directions equivalent focal length) ; \( u_{0i}, v_{0i} \) are the coordinates of the image center (the intersection of the optical axis and image plane).

The relationship between the camera and world coordinate system as:
\[
(X_{\text{st}}, Y_{\text{st}}, Z_{\text{st}}) = \begin{bmatrix}
R & T & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
X_{\text{wt}}, Y_{\text{wt}}, Z_{\text{wt}} \\
1
\end{bmatrix}
\]

(2)

Where, \(X_{\text{wt}}, Y_{\text{wt}}, Z_{\text{wt}}\) are the three-dimensional coordinates of the target in the world coordinate system. The relationship between the image coordinate system and the world coordinate system with can be calculated by equation (3):

\[
\begin{bmatrix}
z_{\text{ci}} \\
v_{\text{ci}} \\
1
\end{bmatrix} = \begin{bmatrix}
f_{\text{ci}} & 0 & 0 & 0 \\
f_{\text{ci}} & 0 & 1 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
X_{\text{st}} \\
Y_{\text{st}} \\
Z_{\text{st}} \\
1
\end{bmatrix} + \begin{bmatrix}
m_{11} & m_{12} & m_{13} & m_{14} \\
m_{21} & m_{22} & m_{23} & m_{24} \\
m_{31} & m_{32} & m_{33} & m_{34}
\end{bmatrix}
\begin{bmatrix}
X_{\text{wt}} \\
Y_{\text{wt}} \\
Z_{\text{wt}} \\
1
\end{bmatrix}
\]

(3)

In Figure 3, point T to be observed has been detected on the two cameras C1 and C2 image. They are separately T1 and T2. Namely T1 and T2 are the same corresponding points T in space. And then

\[
\begin{bmatrix}
z_{1i} \\
v_{1i} \\
1
\end{bmatrix} = \begin{bmatrix}
m_{11i} & m_{12i} & m_{13i} & m_{14i} \\
m_{21i} & m_{22i} & m_{23i} & m_{24i} \\
m_{31i} & m_{32i} & m_{33i} & m_{34i}
\end{bmatrix}
\begin{bmatrix}
X_{\text{st}} \\
Y_{\text{st}} \\
Z_{\text{st}} \\
1
\end{bmatrix} + \begin{bmatrix}
m_{11i} & m_{12i} & m_{13i} & m_{14i} \\
m_{21i} & m_{22i} & m_{23i} & m_{24i} \\
m_{31i} & m_{32i} & m_{33i} & m_{34i}
\end{bmatrix}
\begin{bmatrix}
X_{\text{wt}} \\
Y_{\text{wt}} \\
Z_{\text{wt}} \\
1
\end{bmatrix}
\]

(4)

Eliminating \(Z_{c1}\) and \(Z_{c2}\) in the equation, then get four linear equations about X, Y, Z. Known from analytic geometry, spatial point T is the intersection of \(O_{T1}\) and \(O_{T2}\), it must satisfy the equations:

\[
u_i = \frac{m_{14i} + m_{13i}Y + m_{12i}Z}{m_{13i} + m_{12i}Y + m_{11i}Z}, \quad \nu_i = \frac{m_{24i} + m_{23i}Y + m_{22i}Z}{m_{23i} + m_{22i}Y + m_{21i}Z}, \quad \nu_i = \frac{m_{34i} + m_{33i}Y + m_{32i}Z}{m_{33i} + m_{32i}Y + m_{31i}Z}, \quad \nu_i = \frac{m_{44i} + m_{43i}Y + m_{42i}Z}{m_{43i} + m_{42i}Y + m_{41i}Z}
\]

(5)

The above equations can be abbreviated as:

\[
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
u_4
\end{bmatrix} = h(X, Y, Z)
\]

(6)

4. Robot state estimation based on UKF

4.1. UT transformation

UT transformation principle: Firstly, we take some points by a rule in the distribution of the original state. These points’ mean and covariance are equal to the original mean and covariance. And then putting these points into a nonlinear function is to get the value of the nonlinear function, which form set of points. Using these points is to strike a transformed mean and covariance. Function value got does not require linearization and without ignoring the higher order terms, which are the main points of the UKF better than EKF. Transformation steps are as follows:

Calculation of \((2n + 1)\) Sigma Point

\[
X^{(0)} = \bar{X}, \quad X^{(i)} = \bar{X} + \sqrt{(n + \lambda)}P, \quad i = 1, 2, ..., n, \quad X^{(i)} = \bar{X} - \sqrt{(n + \lambda)}P, \quad i = n + 1, n + 2, ..., 2n
\]

(7)

And then calculate the corresponding weights of these sampling points:

\[
W^{(0)} = \lambda / (n + \lambda), \quad W^{(i)} = \lambda / (n + \lambda) + (1 - \alpha^2 + \beta)
\]

(8)

The parameter \(\lambda\) is a scaling parameter, as:

\[
\lambda = \alpha^2(n + \kappa) - \kappa, \quad \alpha, \beta, \kappa \quad \text{is Positive constants.}
\]

Each sigma point is passed by using the following non-linear formula:

\[
y^{(i)} = g(x^{(i)}), \quad i = 0, 1, ..., 2n
\]

(9)

Calculating the mean and variance of \(y\)

\[
\mu_y = \sum_{i=0}^{2n} W^{(i)}y^{(i)}
\]

(10)
Estimating covariance of $x$ and $y$

$$S_c = \sum_{i=1}^{n} W_i' (y_i - \mu_c) (y_i - \mu_c)^T \tag{11}$$

$$C_c = \sum_{i=1}^{n} W_i' (x_i - m) (y_i - \mu_c)^T \tag{12}$$

$P$ is a positive definite matrix, which is expressed as

$$A = \sqrt{P}, \quad P = AA^T$$

UT can be expressed by using the following matrix form, Sigma point extensions are written in matrix form:

$$X = \begin{bmatrix} \bar{X} & \ldots & \bar{X} \end{bmatrix} + \sqrt{c} \begin{bmatrix} 0 & \sqrt{P} & -\sqrt{P} \end{bmatrix} \tag{13}$$

Sampling point was passed through the nonlinear equation $g(\cdot)$:

$$Y = g(X) \tag{14}$$

And then the Sigma set of points disseminated are weighted and get their mean and covariance as follows:

$$\mu_c = Yw_m, \quad S_c = YWY^T, \quad C_c = XWy^T \tag{15}$$

where, $X$ is the sigma point matrix. Function $g(.)$ calculates $c = \alpha^2 (n+\kappa)$ for each column of $X$, vectors $w_m$ and matrices $W$ can be expressed as:

$$w_m = \begin{bmatrix} W^{(1)}_m \ldots W^{(2a)}_m \end{bmatrix}, \quad W = (I - w_m \ldots w_n) \times \text{diag}(W^{(1)}_m \ldots W^{(2a)}_m) \times (I - w_m \ldots w_n)^T$$

4.2. UKF algorithm applied to state estimation of the target

Based on UT change and then according to Kalman filter principle, UKF puts up state estimation. According to equations (6) and (26), we can obtain simplified form of the system state equation and observation equation:

$$x_k = f(x_{k-1}, k-1) + q_{k-1} \tag{16}$$

$$z_k = h(x_k, k) + r_{k-1} \tag{17}$$

Among them, $x_k \in \mathbb{R}^n$ is the state of time $k$, $z_k \in \mathbb{R}^m$ is the observed value of time $k$. $q_{k-1} \sim N(0, Q_{k-1})$ is a Gaussian process noise. $\eta \sim N(0, R_k)$ is Gaussian observation noise.

$$x_k = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} \end{bmatrix}, \quad z_k = \begin{bmatrix} u_{1k} & v_{1k} & u_{2k} & v_{2k} \end{bmatrix}$$

Among them, $u_{1k}, v_{1k}$ are the image coordinates of the target at time $k$ in the upper arm camera. $u_{2k}, v_{2k}$ are the image coordinates of the target at time $k$ in the under arm camera.

According to the matrix form of UT, UKF is expressed as the following steps. Initialization

$$\hat{x}_0 = E[x_0], \quad P_0 = E[(x - \hat{x}_0)(x - \hat{x}_0)^T] \tag{18}$$

Forecast: calculate the predicted state mean $\hat{x}_k$ and predicted covariance $\hat{P}_k$.

Firstly, according to UT, sampling points are got at k-1 time.

$$x_{k-1} = \begin{bmatrix} x_{k-1} \ldots x_{k-1} \end{bmatrix} + \sqrt{c} \begin{bmatrix} 0 & \sqrt{P_{k-1,1}} & -\sqrt{P_{k-1,1}} \end{bmatrix} \tag{19}$$

Putting the transformed sampling points into the state of the nonlinear equation of demand at time k:

$$x_k = f(x_{k-1}, k-1) \tag{20}$$

So the mean and covariance are got at time $k$

$$\hat{x}_k = x_k w_m, \quad \hat{P}_k = x_k W x_k^T + Q_{k-1} \tag{21}$$

Update: calculate the covariance of the forecast mean and variance of observation and state and covariance observation

First update the sampling point

$$\hat{X}_k = \begin{bmatrix} \hat{x}_1 \ldots \hat{x}_1 \end{bmatrix} + \sqrt{c} \begin{bmatrix} 0 & \sqrt{P_{k}} & -\sqrt{P_{k}} \end{bmatrix}$$

Double-Robot Dynamic Stereo Vision Algorithm based on UKF and its Application in Moving Target Tracking

Huiying Dong, Xue Lu, Yuqing He, Shengfu Chen, Jianda Han
Predict observations at time k:

\[ \hat{z}_k = h(\hat{X}_k, k) \]  

Final calculate the mean and covariance of observations \( \mu_k \) and \( S_k \) as well as the covariance between sampling points and observations \( C_z \):

\[ \mu_k = \hat{z}_k W_m, \quad S_k = \hat{z}_k W [\hat{z}_k]^T + R_k, \]  

(4) Calculate the filter gain \( K_k \), at the same time updated state mean \( \hat{X}_k \) and covariance \( P_k \):

\[ K_k = C_z S_k^{-1}, \quad x_k = \hat{x}_k + K_k [z_k - \mu_k], \quad P_k = \hat{P}_k - K_k S_k K_k^T \]  

5. Collaborative tracking

Collaborative tracking treats the observed target as a Leader, and treats other tracking UAV as a follower. Each Follower tracks the target' position and attitude information according to fusion algorithm to keep track of the formation which uses the method mentioned in literature [14]. Collaborative tracking structure is shown in Figure 4.

[Figure 4. Control Structure of Collaborative Tracking]

where f, l, h stands for the three directions x, y, and z coordinates and u, v, w stands for x, y, and z-direction speed, p, q, r stands for the roll rate, pitch rate, yaw rate separately, uref, vref and wref are control parameters, \( \psi, \sigma \) is the deviation angle.

6. Experimental platform and experimental results

6.1. Experimental platform

6.1.1. Experimental platform

Experiment’s observation platform of cooperation is shown in Figure 5. The system consists of a three-arm spindle: one end of the upper arm install aircraft, the other is the CCD 1; one end of the middle arm install plane, the other end is the to-be-observed target ‘Target’; one end of the lower arm install the aircraft, the other end is the CCD 2. Each arm has three degrees of freedom, that the level of freedom, the vertical degree of freedom and attitude degrees of freedom. Each UAV can autonomous fly without collision between the UAV through the physical limit. The 850 nm LED is the to-be-observation, and the 850 nm filter is installed on each CCD, which has good noise immunity and can filter out the background having nothing to do with the target, with good usability.

The position and orientation of the observing robot (middle arm observation target) is obtained from a variational three-dimensional vision which is consist of the above arm and lower arm CCD.
6.1.2. Calibration parameters of the camera

Assume that UAV-i’s vertical direction rotation angle is $\alpha_i$ and rotation angle of the horizontal direction is $\beta_i$ at time $t$ (“the distance between aircraft and axis is fixed, the aircraft's three position can be completely confirmed by these two angles”). Therefore the camera coordinate system relative to the world coordinate system’s rotation and translation matrix as well as the world coordinates are as follows:

$$R_w = \begin{bmatrix} \cos \alpha_i \cos \beta_i & \cos \alpha_i \sin \beta_i & \sin \alpha_i \\ -\sin \alpha_i \cos \beta_i & \cos \beta_i & 0 \\ -\sin \alpha_i \sin \beta_i & -\sin \beta_i & \cos \alpha_i \end{bmatrix}, \quad T_w = \begin{bmatrix} l, -h \sin \alpha_i \\ 0 \\ -h \cos \alpha_i \end{bmatrix},$$

$$\begin{cases} x_{w} = -l \cos \alpha_i \cos \beta_i \\ y_{w} = -l \cos \alpha_i \sin \beta_i \\ z_{w} = h_l - l \sin \alpha_i \end{cases}$$

UAV’s $R_w$ and $T_w$ can be obtained according to $\alpha_i$ and $\beta_i$ returned by the encoder sensor every moment. Zhang Zhengyou calibration method is adopted to calibrate the camera internal parameters. They are as follows:

a) The upper arm CCD 1
u0=344.8896; v0=314.5257; fx=613.79111; fy=673.76303;  
b) The lower arm CCD 2
u0=336.3556; v0=328.5333; fx=595.23521; fy=650.55831;

6.1.3. State equation

The experiment tracks three-dimensional moving target and the target state equation is:

$$X(k+1) = \Phi X(k) + V(k) \quad (26)$$

where, $\Phi$ is the state transition matrix and $V(k)$ is the Gaussian white noise. The state vector is:

$$X = [x \ y \ z \ \dot{x} \ \dot{y} \ \dot{z}]^T$$

where, $(x, y, z)$ is the three state Department components under the right angles respectively, $(\dot{x}, \dot{y}, \dot{z})$ is the three-direction velocity component.

6.2. Experimental verification

6.2.1. Experimental conditions

Experimental platform is shown in Figure 5, the upper and lower arm CCD always keep the observation state through the establishment of the stereo vision system to observe and follow-up observations targets. Motor process of the to-be-observed target (middle arm) is: UAV initialize the system and maintain the level state during $1s \sim 23s$; target observation maintains level state and rotate horizontal at a constant speed of $5^\circ/s$ during $24s \sim 38s$; maintain the vertical angle of $-15^\circ$ for target observation and rotate horizontal at a constant speed of $5^\circ/s$ during $39s \sim 57s$; target observation maintains level state and rotate horizontal at a constant speed of $5^\circ/s$ during $58s \sim 75s$; maintain the vertical angle of $15^\circ$ for target observation and rotate horizontal at a constant speed of $5^\circ/s$ during $76s \sim 96s$.

Assume that the environmental impact of X, Y, Z direction is the same as the interference of the horizontal acceleration and vertical acceleration are independent Gaussian white noise, observation errors are Gaussian white noise as well. The initial state $x_0 = [-326, -600, 1477.5, 0, 0, 0]^T$ (location unit: mm, speed unit: mm/s), set $\alpha = 1, \ \beta = 2, \ \kappa = 1$ in the UKF arithmetic.
6.2.2. Experimental results and analysis

Figure 6, Figure 8, Figure 10 is the UKF estimation method based on stereo vision comparison of estimation methods based on monocular vision of the UKF in X direction, Y direction, and Z direction respectively. The solid line is the actual value, the dashed lines are the UKF estimates based on stereo vision; point type lines are the UKF estimated value based on monocular vision. Figure 7, Figure 9, Figure 11 is the error of stereo vision and monocular vision in X direction, Y direction, Z-direction respectively. Solid line is the estimation error of the UKF based on stereo vision; the dashed line is the estimation error of the UKF based on monocular vision.

Table 1. Comparing of error EMR (Unit: m)

<table>
<thead>
<tr>
<th>arithmetic</th>
<th>X direction</th>
<th>Y direction</th>
<th>Z direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>monocular vision</td>
<td>0.06816</td>
<td>0.09946</td>
<td>0.09253</td>
</tr>
<tr>
<td>Dynamic stereo vision</td>
<td>0.03641</td>
<td>0.05324</td>
<td>0.05565</td>
</tr>
<tr>
<td>stereo vision</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6, Figure 8 and Figure 10 shows that monocular vision and dynamic stereo vision can be a good predictor of the target motion state and can follow the observation target well; Figure 7, Figure 9, Figure 11 and Table 1 shows that the UKF estimation method based on dynamic three-dimensional visual’s estimation accuracy is higher than the UKF estimated based on monocular vision, and dynamic three-dimensional visual tracing’s accuracy significantly is better than monocular vision-based tracking method.

7. Conclusions

The UKF method is used in multi-robot system. According to the principle of stereoscopic vision, establish dynamic stereo vision system and fuse the two aerial robot observations. Experimental results show the UKF estimation method based on dynamic stereo vision significantly better than that based on monocular vision tracking method. The resulting is that integration of observations of different robots can improve the accuracy and completeness of context-aware, getting more accurate information than a single robot.
8. References


