Dynamical Vision Sensors based Active Cooperative Observation in Three Dimensional Environment

Feng Gu Member, IEEE, Yuqing He Member, IEEE and Jianda Han Member, IEEE

Abstract — Vision based observation of a moving target is one of the important problems and hot issues of mobile robot system. Dynamical vision sensors which is constituted with multiple robots to obtain improved observational results has been shown to be a good substitution of single vision sensors. Thus, a new active cooperative observation (ACO) method based on two dynamical vision sensors is proposed. The most characteristics of the proposed method is that data fusion and path planning algorithm are simultaneously implemented and combined with each other by the optimal observation formation so that the influence of relative positions among MVSs and target on the cooperative observation result can be fully considered to improve the observation result. Finally, for verifying the validity and feasibility of the proposed method, experiments on a multiple rotor flying robots test-bed are conducted and analyzed, respectively.

Key words: dynamical monocular vision, active cooperative observation, set-membership filter

I. INTRODUCTION

Vision based Moving Target Observation (VMTO) has been utilized in many applications, such as, suspicious character detection in border patrol, friend robot localization in multiple robot system, etc[1][4]. For some cases that target moves in 2D space, monocular vision is enough to implement the estimation and tracking of the moving target. However, in most other cases, high and real time 3D VMTO is necessary. One of the key problems in 3D VMTO is to obtain the 3D information of the target by data fusion from different cameras. However, the existing method such as Depth From Focus (DFF) [5], Depth From Defocus (DFD) [6] may present acceptable accuracy only in some special cases since they have very small physical size [7]. Comparatively, stereo vision usually has much larger physical size (i.e., baseline) and thus is more accurate for the 3D applications and more often used in real applications of mobile robot systems. For example, in reference [8] and [9], it is used to detect the dynamic target and to model the 3D terrain map of the ground environment, respectively. Unfortunately, the high precision of stereo vision system is mainly due to the large baseline [7], while the physical sizes of most robot systems make it impossible to increase it as desired. This will be much heavier in the cases when the distance between the sensor and the target is much larger than the baseline.

Based on the problem above, some effective substitute strategies have been developed to obtain accurate observational results by constructing a system composed of multiple mobile robots which are equipped with several monocular or stereo cameras. However, most of the researches focus on how to improve the observational precise by using data fusion algorithm. In reference [10], re-parameterization of 2D Gaussian distributions fusion method is utilized to improve the observation accuracy by combining information from more than two cameras; The authors of reference [11] research the simultaneous localization and mapping (SLAM) problem by using two cooperative single monocular vision sensors, the visual data from which are treated by monocular methods and fused by the SLAM filter; in [12], Bayes estimator based algorithm for cooperative localization is proposed by fusing the bearing-only information provided by multiple cameras. Another strategy to improve the observation results using multiple robots is to regulate the coordinate behavior of the robots. Some researchers have proved that the relative position will influence the fused results, which is more obviously when both the sensor and the target are movable. Besides that, the occlusion is unavoidable and will deteriorate the observational results heavily in complicated environment, especially in dynamical environment. Thus, the path planning method should be involved to predict the possible occurrence of occlusion and then avoid it through regulating the behavior of mobile robot, or vision sensor. However, up to now, there are little work about coordination behavior planning method aiming at improve the 3D VMTO problems.

In this paper, we try to research the problem of simultaneously data fusion and coordination behavior optimization of two movable monocular vision sensors aiming at moving target observation, i.e., active cooperative observation (ACO). This problem may involve the following two sub-problems: 1) data fusion algorithm; 2) coordination behavior optimization algorithm. For the first sub-problem, the data fusion is achieved with the more robust Set Theory based Estimation (STE) method which assumes that the measurement noise and process noise are both Unknown But Bounded (UBB) [13] and can produce an estimated uncertain set which the real system state are ensured in[14]. For the second sub-problem, a relative velocity coordinates (RVCs) based planning algorithm is introduced into the data fusion algorithm. In the new proposed method, the coordinate behavior of the robots is regulated based on the data fusion results through a concept of optimal observational condition. Also, the occlusion problem can also be avoided utilizing the...
concept of obstacle avoidance.

II. PROBLEM DESCRIPTION
The sketch of 3D VMTO using two robots equipped with monocular vision sensor (MVS) is shown in Fig.1, where both the target and the MVSs can freely move in 3D space.

Fig. 1 Sketch of the 3D VMTO using two movable MVS

* C1 and C2 are the image planes of two monocular vision systems; T1 and T2 are the image points of target T in the image plane.

Based on the basic principle of stereo vision, the 3D location of the target can be obtained by combing two cameras’ measurements with the rotation matrix and translation vector, R and T, between camera and world coordinate system. However, in this paper, both of the two monocular vision sensors are movable. That means the relative pose of them is not fixed and highly relates to the posture of the MVS, i.e., [R, T] should be rewritten as,

\[ \begin{bmatrix} R(x, T(x)) \end{bmatrix} \] (1)

Since the target is also movable, the estimation of it can be denoted as the following estimation problem,

\[ x_{T,k+1} = f(x_{T,k}, w_{T,k}) \] (2)

\[ y_i = h(x_{1,k}, x_{T,k}, n_i) \] (3)

where Eq. (2) is the motion equation of the moving target; Eq. (3) model the measurements of two dynamical monocular vision sensors; \( x_{1,k} \) and \( x_{T,k} \) (i = 1, 2) are the state vector of target and MVS, respectively; \( w_{T,k} \) is the model uncertainty; and \( n_i \) is measurement errors; subscript \( k \) and \( k+1 \) means the corresponding variable is at time instant \( k \) and \( k+1 \); \( y_i \) is the measurement vector of each vision sensor, i.e.,

\[ h(x_{1,k}, x_{T,k}, n_i) = (u_i, v_i, u_2, v_2)^T + n_i \] (4)

where, \( u_i, v_i \) are the pixel coordinates measured by MVS; Then the target states can be estimated by using system equation Eq. (2) and measurement equation Eq. (4) with some estimation algorithm, which can be denoted as a map as follows,

\[ x_{T,k+1} = Z(x_{T,k}, y_i, w_{T,k}, n_i) \] (5)

**Remark 1:** In general, the detailed motion equation of the target is difficult to be obtained. In order to predict the motion of the target, the noise-driven integral or double integral equation is usually used to model the target, i.e.,

\[ \begin{cases} X_{T,k+1} = X_{T,k} + v_{T,k} \cdot \Delta T + w_{1,k} \\ v_{T,k+1} = w_{2,k} \end{cases} \] (6)

where \( X_{T,k} = (x_{T,k}, y_{T,k}, z_{T,k})^T \) and \( v_{T,k} = (v_{x,k}, v_{y,k}, v_{z,k}) \) are the position and velocity of the target at step \( k \) respectively; \( \Delta T \) is sampling time; \( w_{1,k} \) and \( w_{2,k} \) is the process error.

**Remark 2:** Compared to the static stereo vision system, one of the most disadvantages of the movable stereo system proposed in this paper is that the relative posture of different vision sensors is not fixed and obtainable with super high precision. These errors are introduced into the system by the states of each MVS system, i.e., [R, T] should be rewritten as,

\[ y_i = h(x_{1,k} + e_{1,k}, x_{2,k} + e_{2,k}, x_{T,k}, n_i) \] (7)

In order to avoid dealing with the extra errors and thus complicating the observation problem, some pre-processing can be introduced, i.e., linearization, to include errors \( e_{1,k} \) and \( e_{2,k} \) into the errors \( n_i \). That means, when we model \( u_i \), we should consider not only the measurement error from vision sensors but also the measurement errors from the state of each MVS system.

Another problem is to make full use of the dynamics performance of the MVS to improve the observation accuracy. The model of MVS system for observation and tracking should also be considered for the path planning. Assume that the motion of MVS is dominated by the following equation,

\[ x_{1,k+1} = f(x_{1,k}, u_{1,k}) \] (8)

where \( f(*) \) are some pre-defined map (linear or nonlinear); \( x_{1,k} \) and \( x_{1,k+1} \) are the states of the \( i \) th MVS at the time instant \( k \) and \( k+1 \), respectively. In this paper, the coordinate behavior planning means to find some mapping which connect the current states of both MVSs and target with the optimal behavior in the sense of some pre-defined cost function, which includes the data fusion results (the so-called optimal observation formation in section III). This problem can be modeled as constructing explicit or implicit mapping \( \Xi \) as follows,

\[ u_{k, optimal} = \Xi(x_{1,k}, x_{2,k}, x_{T,k}, y_i) \] (9)

where \( x_{T,k} \) is the predicted state of the moving target at time instant \( k \).

**Remark 3:** In this paper, the following kinematic equation is used to denote the action of MVS system and moving target for the common applications,

\[ \begin{cases} x_{1, k+1} = x_{1, k} + v_{1, k} \cdot \Delta T + u_{1, k} \\ y_{1, k} = y_{1, k} + v_{y, k} \cdot \Delta T \\ \Delta(x_{1, k}, y_{1, k}) \leq 0 \end{cases} \] (10)

Thus, ACO problem can be described as follows: With system and measurement equation (Eq. (2) - Eq. (4)), design the data fusion estimation algorithm \( Z(*) \) to achieve 3D target state estimation with two monocular vision sensor. Then the coordinate behavior of two mobile MVSs is planned aiming at better data fusion results. The whole process and main idea of ACO is sketched in Fig. 2.
III. ACTIVATION COOPERATIVE OBSERVATION ALGORITHM

In this section, the optimal observation formation which is the key link between data fusion and path planning is first proved. Then data fusion algorithm and path planning strategy will be introduced.

A. Optimal observation formation

In order to explain and compute the optimal observation formation, we’d like to firstly introduce the procedure of monocural vision observation as shown in Fig. 3.

In the left figure of Fig. 3, the real position of target is denoted as a ‘*’, and then based on the UBB condition as in the last section, each monocular vision sensor can obtain a measurement ellipsoid, which is embedded in the image plane and circle the real position of the target. Because the observation is two-dimensional, the real observation uncertain set in 3D space can be denoted as a cylinder whose axis is perpendicular to the corresponding image plane. Thus, the data fusion result can be denoted as the intersection of the two cylinders smallest through regulating the relative position among the target and MVSs, i.e., the relative posture of the two cylinders.

![Fig. 3 Relative relationship between two observation uncertainty set](image)

(a) sketch of the data fusion  (b) relative posture of two cylinders

Two variables will influence the intersection of the two cylinders: one is the relative position between them, i.e., the smallest distance of the central axes of them; the other is the relative attitude between them, i.e., the angle between the axis of them. The relative attitude can be easily changed through regulating the relative position of two MVSs. However, the relative position can not be defined beforehand because the measurement result of each monocular vision sensor are stochastic distributed in an directed ellipsoid, thus our so-called optimal observation formation can only be denoted as the relative attitude of the two cylinder and thus can be computed by solving the following min-max optimization problem,

$$\min \max V(C_1 \cap C_2)$$

(12)

where $C_1$ and $C_2$ denotes the two cylinders, respectively, as shown in Fig. 3; $V(*)$ means the volume of *; $a$ and $\alpha$ are the smallest distance and the angle between the central axes of $C_1$ and $C_2$.

The following theorem about optimal observation formation is not difficult to be proved.

**Theorem 1:** Assumption that there is no difference of the observation uncertainty for each monocular vision sensor at the two different directions in the image plane, the optimal observation condition, i.e., the solution of problem(12), is $\alpha = \pi/2$.

**Proof:**

The condition of Theorem means the two 2D measurement uncertainty sets are two circles. This is usually reasonable since the difference between two directions is very small.

Theorem I can be shown using the following two steps: 1) the volume of intersection between the two cylinders is maximum when $a = 0$ regardless of the angle $\alpha$; 2) the optimal angle $\alpha$ is $\pi/2$ when $a = 0$.

The first step can be directly obtained in reference [16]. Now we will show the second step.

Without considering the absolute posture of the two cylinders, the following equation can be used to denote them,

$$x^2 + y^2 \leq r_1^2$$

(13)

$$x^2 + (y \cos \alpha + z \sin \alpha)^2 \leq r_2^2$$

(14)

Thus, the intersection between them can be computed using the following double integration in the area denoted as Eq. (13). Described in the cylinder coordination, the intersection can be denoted as,

$$\frac{2}{\sin \alpha} \int_0^{2\pi} \int_0^{r_1} \sqrt{r_2^2 - r^2 \sin^2 \theta} dr d\theta$$

(15)

where $r$ and $\theta$ denote the coordinates in cylinder coordinate of 3D space.

From Eq. (15), it is easy to conclude that the volume of the intersection is smallest if $\alpha = \pi/2$. This completes the proof of Theorem 1.

B. ESMF based cooperative observation

In this section, we will introduce how to cooperatively estimate the 3D position of the target using the system equation (2) and measurement equation (4) based on the set theory based filter, i.e., Extended Set-Member Filter (ESMF). ESMF is nonlinear guarantee estimation and filter method which supposes the noise is unknown but bounded (UBB) and often modeled as an ellipsoid [13]. Similarly, the estimated states are also denoted as some ellipsoid set given by the following equation [14],

$$E(\hat{x}, P) = \{x \in \mathbb{R}^n | (x - \hat{x})^T P^{-1} (x - \hat{x}) \leq 1\}$$

(16)

where $\hat{x}$ is the center of the ellipsoid; $P$ is an envelope matrix satisfying symmetric and positive definite conditions.
Similar to Kalman class filter, ESMF involves also two steps as follows and Fig. 4:

- **Prediction**: to predict the system state in next instant according to the previous state and system model, the prediction result is a prediction set that contains the real state of the target;
- **Update**: to revise the prediction using the measurement.

With the observation problem model (Eq. (2) and (4)) the state of the target can be estimated using ESMF whose detail can be found in [13]. Then the 3D position of the target can be obtained with two monocular visions by cooperation.

**Remark 4**: From the preceding procedure, it is not difficult to understand why the ESMF algorithm is more robust than Kalman class filter. This is mainly because the former is easy to consider all kinds of errors and noises and ensure the real value is in the estimated uncertainty set. Of course, this procedure maybe introduces some certain of conservativeness. This problem, fortunately, can be alleviated by good data fusion strategy as that in this paper.

### C. Path planning for active cooperative observation

The implementation of ACO can be shown as Fig. 5, where the path planning algorithm is embedded into the cooperative observation between prediction step and update step as discussed in Section III so that the MVSs can plan their trajectory according to the prediction result to satisfy the optimal observation formation and then achieve data fusion.

**Remark 4**: From the preceding procedure, it is not difficult to understand why the ESMF algorithm is more robust than Kalman class filter. This is mainly because the former is easy to consider all kinds of errors and noises and ensure the real value is in the estimated uncertainty set. Of course, this procedure maybe introduces some certain of conservativeness. This problem, fortunately, can be alleviated by good data fusion strategy as that in this paper.

The path planning for ACO should consider the following factors: 1) **optimal observation formation**, 2) **target pursuit**, 3) **obstacle avoidance**, and 4) **collision avoidance**. With LP based RVCs framework above, the path planning for ACO problem is solved with the following steps:

- a) Build each factor’s function related to the each MVS’s relative velocity \( v_{\text{rel}} \) according to the predictive result and MVS model, i.e. \( F = f(v_{\text{rel}}) \), where \( F \) is quantization parameters of each factor;
- b) Linearize \( f(v_{\text{rel}}) \) and building the optimization problem with LP method with factor 1) as optimal objective and other factors as constraint condition;
- c) Solve the optimal problem in b) and obtain the output of the required velocity \( v_i \).

### IV. EXPERIMENTS AND ANALYSIS

#### A. System setup

In order to verify the proposed algorithm, a demonstration experiment has been conducted with respect to the multiple rotor-flying-robots (MRFRS) testbed, which is a new-designed platform to verify the cooperation algorithms of multiple-flying-robot and introduced in detail in [18].

Experimental platform is shown in Fig. 7, which is composed of three arms with one side fixed and the other side equipping a small rotor flying robot. In this experiment, the middle manually controlled robot (defined as R1) is used as a moving target whose motion state is unknown for the other robots, and the other two robots (defined as R2 and R3, respectively)
are taken as two MVSs required to cooperatively observation R1. It should be noted that the motion of the target (R1) as well as two MVSs (R2, R3) is directly measured with the encoder equipped in each rotary joint and the measurement result can be described as vertical angle \( \alpha_i \) and horizontal angle \( \beta_i \) as shown in Fig. 7-a. Due to the limitation of the arm, which is described as Eq.(17), the motion of moving target cannot move completely freely. However, we only use this limitation to compute the 3D position of the target and MVSs as the referenced real value, while during the 3D localization experiments, the 3D motion equation, as discussed in Section II, Eq.(2) and(4) are still used.

\[
\begin{align*}
\dot{x}_r &= -l_1 \cos \alpha \cos \beta_i \\
\dot{y}_r &= -l_1 \cos \alpha \sin \beta_i \\
\dot{z}_r &= h_1 - l_1 \sin \alpha,
\end{align*}
\]

(17)

where, \( h_1 \) can be found in Fig. 7-a.

The envelope matrix of system noise and measurement noise in Eq.(2) and (4); initial envelope matrix \( P_i \) are as follows respectively:

\[
Q = \begin{bmatrix} 20 & 5 & 6 \\ 5 & 20 & 7 \\ 6 & 7 & 30 \end{bmatrix}, \quad R = \begin{bmatrix} 8 & 1 & 1 & 1 \\ 1 & 10 & 1 & 1 \\ 1 & 1 & 8 & 1 \\ 1 & 1 & 1 & 10 \end{bmatrix}, \quad P_i = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 100 & 0 \\ 0 & 0 & 100 \end{bmatrix}
\]

B. Experiment result and analysis

Because of the limitation of the testbed, the target is controlled to track the following desired trajectory,

\[
\begin{align*}
\dot{x}_r &= -l_1 \cos \alpha \cos \beta_i \\
\dot{y}_r &= -l_1 \cos \alpha \sin \beta_i \\
\dot{z}_r &= h_1 - l_1 \sin \alpha,
\end{align*}
\]

(18)

where,

\[
\alpha_i = 0^o, \quad \beta_i = \begin{cases} 
\beta_i & t \leq 20s \\
5t & 20s < t \leq 95s \\
475 + 3t & 95s \leq t
\end{cases}
\]

(19)

\( \beta_i \) is the initial horizontal angle of the target as shown in Fig. 7. Before analyzing the experiment result, the following points should be explained first for understanding the result better.

1) For the target the position that calculated with Eq. (17) according to the encoder measurement \((\alpha_i, \beta_i)\) is taken as the referenced real state of the target.

2) According to the pose parameters of the cameras equipped in the testbed, it can be calculated that when the horizon angle between the arm of R2 and that of R3 is 110°, the cooperative observation angle equals to 90 degree, i.e., the optimal observation condition is satisfied and the intersection will have the smallest area. That means, the observation formation in the motion planning should be,

\[
|\beta_i(k) - \beta_i(k)| = 110^o
\]

(20)

3) In the testbed due to the limitation of the arm, the motion of the target is actually a 2D motion and the angles \((\alpha_i, \beta_i)\) can be totally measured even with single CCD. Thus, the 3D coordinate \((x_r, y_r, z_r)\) of the target can also be obtained with single CCD by converting \((\alpha_i, \beta_i)\) to \((x_r, y_r, z_r)\) using Eq.(17).

4) As discussed in Section III, the estimation result is the uncertainty set (expressed as an ellipsoid) that described as Eq.(16). Then define that,

\[
F \triangleq [x - \hat{x}]^T P^{-1} [x - \hat{x}]
\]

(21)

where, \( \hat{x} \) is the center of the ellipsoid; \( P \) is the symmetrical and positive defined matrix. For the point \( x \) that makes 0\( \leq F \leq 1 \), it indicates that \( x \) lie within the ellipsoid \( E(\hat{x}, P) \). In our experiment, we use this relationship to check whether the real states of the target lie with the estimated uncertainty set. Thus, \( F \) in Eq. (21) can be taken as the quantitative index of the reliability.

Experimental results are as shown in Fig. 8-Fig. 12. Fig. 8 is the overview of the cooperative observation procedure.

![Fig. 8 Overview of the cooperative observation result](image)

*The ellipsoids are the estimated uncertainty sets that contain the real state of the target.

Fig. 9 gives out the value of \( F \) in Eq. (21) at each time instant. From the figure it can be seen that all the results are in the interval (0, 1), that means all the true value point of the target \( x_t \) are lie within the uncertainty set as shown in Fig. 8.

![Fig. 9 Reliability index](image)

From Fig. 8 and Fig. 9, the following conclusion can be obtained: the proposed method can make sure the true state of the target lie within the observation result, i.e. the uncertainty set, which verifies the reliability of the method.

![Fig. 10 Cooperative observation result](image)
We have discussed in Section III that the ESMF based observation result is an uncertainty set that contains the real state of the target. Thus, the size of the uncertainty is also an index of the observation accuracy. In the experiment, the size of uncertainty set between single and cooperative observation has been compared to show the validity of the proposed method. (The reason why the single observation result can be obtained had been explained in point 3) in this section). In Fig. 11, the dashed and solid lines are the single and cooperative observation result respectively. From the comparison, the uncertainty size of cooperative observation result is much smaller and smoother than that of single one, which indicates the proposed method has better accuracy and robust.

![Fig. 11 Comparison between single and cooperative observation](image1)

Fig. 11 Comparison between single and cooperative observation

Fig. 12 shows the cooperative angle during the experiment. It can be seen that the cooperative angle can be kept around the expected angle and achieve optimal cooperative observation with the proposed ACO method.

![Fig. 12 Cooperative angle](image2)

Fig. 12 Cooperative angle

V. CONCLUSION

In this paper, the problem of moving target observation in 3D space using only two mobile monocular visions is researched, and a so called active cooperative observation (ACO) algorithm is proposed. The main original contributions of this paper are in the following points: 1) A new framework of ACO is proposed, where the data fusion and path planning are connected to each other by the concept of optimal observation formation. By doing this, multiple measurements’ fusion result can be improved further through regulating the relative posture among mobile sensors and target. 2) Furthermore, with respect the interested problem in this paper, i.e., ACO of two mobile monocular vision sensors, the optimal observation formation, denoted by an angle, is given and shown in detail. 3) An Extended Set-Membership Filter based data fusion algorithm and a relative velocity based path planning algorithm is proposed to implement the ACO algorithm.

Moreover, the experiments with respect to a multiple rotor flying robots test-bed are conducted. Through analyzing the experimental data and results, the feasibility and validity of the proposed methods in the application of moving target observation using only two mobile monocular vision sensors have been verified. Finally, It should be noted that the proposed methods is not limited in the application of multiple monocular vision system, it can also be used in other multiple-mobile-sensor-system (MMSS), or even heterogeneous MMSS.

REFERENCES


