Multi-Robot Cooperative Stereo for Outdoor Scenarios

A. Dias, J. Almeida, E. Silva
INESC TEC Robotics Unit
School of Engineering Polytechnic Institute of Porto
Email: adias, jma, eaps@lsa.isep.ipp.pt

P. Lima
Institute for Systems and Robotics
Instituto Superior Técnico (IST), Lisbon Portugal
Email: pal@isr.ist.utl.pt

***Abstract*—In this paper, we propose a cooperative perception framework for multi-robot real-time 3D high dynamic target estimation in outdoor scenarios based on monocular camera available on each robot. The relative position and orientation between robots establishes a flexible and dynamic stereo baseline. Overlap views subject to geometric constraints emerged from the stereo formulation, which allowed us to obtain a decentralized cooperative perception layer. Epipolar constraints related to the global frame are applied both in image feature matching and to feature searching and detection optimization in the image processing of robots with low computational capabilities. In contrast to classic stereo, the proposed framework considers all sources of uncertainty (in localization, attitude and image detection from both robots) in the determination of the objects best 3D localization and its uncertainty. The proposed framework can be later integrated in a decentralized data fusion (DDF) multi-target tracking approach where it can contribute to reduce *rumor propagation* data association and track initialization issues. We demonstrate the advantages of this approach in real outdoor scenario. This is done by comparing a stereo rigid baseline standalone target tracking with the proposed multi-robot cooperative stereo between a micro aerial vehicle (MAV) and an autonomous ground vehicle (AGV).**

I. Introduction

In recent years, we have seen a growing research effort on novel multi-robot cooperative tasks for heterogeneous mobile robotics applications. This ongoing development is driven by a significant number of potential end-user applications, where is necessary to reduce the human in the loop interaction which includes large-scale sensing operations[12], cooperative search and rescue tasks[5], surveillance[7], recognition reconnaissance and border control[8]. Currently, mobile robots employed on these high-end user applications are equipped with state-of-the art sensing equipment allowing them to navigate and perceive their surrounding environment. One of the most common and versatile means of perception in mobile robotics applications is visual sensing with one or more cameras which are able to acquire visual information[13] based on cooperative approaches. Taking this a step further, here we address an outdoor multi-robot scenario without localization issues, with the surveillance task goal of detecting and estimating 3D high dynamic targets positions behavior in a cooperative vision flexible and dynamic stereo baseline framework.

State of art approaches to enumerated end-user applications can be organized according to cooperative tasks emerged from local or cooperative perception.

In local perception approach, each robot is capable of detecting and locating targets, sharing that information over some communication middleware that can be later used to some cooperative mechanism for task allocations[10].

Considering the proposal scenario, those approaches present several limitations in any possible vision setups: monocular or stereo rig baseline. In monocular vision, we have the intrinsic difficulty in estimating depth and absolute scale[1], so 3D target estimation without target known size is a research challenge. Techniques like SFM(struture-from-motion) or SLAM(Simultaneous localization and mapping) are able to estimate depth from a monocular camera[6][3], but the scene must have a large field of view and motion must not occur along the optical axis and preferably parallax motion to allow a fast uncertainty map convergence[1]. SLAM techniques are able to obtain good results in depth estimation for indoor and even in outdoor map building scenarios although with constraints such as high computational requirements (not available in most of the robots with low payload), lower camera dynamic, preferably with features loop closing and large field of view, but unable to track targets with high dynamic behavior. Still with monocular vision and for a particularly case of aerial vehicles depth estimation can be obtained based on flat earth assumption[2]. Although it is simple, its application is limited to tracking objects on the ground with low accuracy and not applied to our addresses scenario. Regarding stereo rig baseline, 3D target estimation is a well known solution due to its relatively simple image scale and depth estimation although with limited application when the goal is to track targets whose depth distance greatly exceeds the available stereo rig baseline, therefore reducing the stereo setup to a bearing-only sensor[15]. The estimation error grows quadratically with the depth[15][4], becoming even more relevant this limitation when the robot majority tends to decrease its scale factor and consequently smaller rig baseline. The enumerated limitations strengthens our proposal by having a multi-robot monocular approach with a flexible and dynamic baseline between robots able estimate 3D information from correlated detected targets.

Focusing now in cooperative perception approaches, characterized by each robot, available at the multi-robot formation, builds its own local partial representation of the world, described by the belief state and share in order to improve their knowledge. Some of this methods are: Decentralized Data Fusion(DDF)[11] by incorporating 2D measures possible to be represented by Gaussian Mixture Model(GMM)[9], Cooperative SLAM[14] and for the special case of indoor scenarios a decentralized EKF monocular camera inertial sensor fusion method[1] to recover the relative configuration between monocular cameras. Common to all enumerated methods is the requirement translation of the information received from other robots to the same local representation. This step is critical in order to avoid *rumor propagation* that could lead to overoptimistic estimations. In Cooperative SLAM[14] this problem was considered and solved thought the epipolar geometric constraint between cameras. This is part of your proposal in which we detail in section II-C.

A. The Aim of this Work

We propose a method to estimate the 3D target information based on multi-robot vision bearing-only measurements in outdoor scenarios.

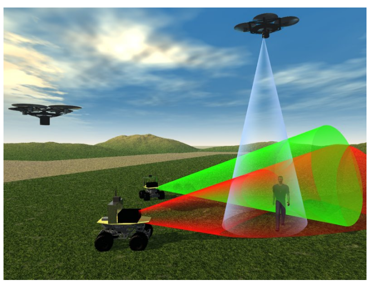

Fig. 1. Multi-Robot Cooperative Stereo

The relative positions and orientations between monocular cameras are allowed to change over time, which means that we are able to form a flexible stereo baseline and establish overlap views based on geometric constraints emerged from the multi-robot collaborative stereo formulation and provide a 3D outdoor localization for multi-targets with high dynamic behavior (see figure 1). The envisioned multi-robot cooperative stereo framework can combine monocular vision information from heterogeneous vision sensors included , but is not limited to, infrared thermographic camera, visible camera and multi-spectral cameras which means that we can have multiple robots cooperating in the same environment and combining the information provided by each vision sensor. As regards DDF target tracking approaches the framework can be applied as layer able to support data association and avoid *rumor propagation* between robots and in the initialization process of new targets.

The paper is organized as follows: in section II we present the multi-robot cooperative stereo framework and detail the developed blocks. Section III describes the outdoor scenario and the vehicles used to obtain the results detailed in section IV, followed by conclusions and future work in Section V.

II. Multi-Robot Cooperative Stereo Framework

The general scheme for the multi-robot cooperative stereo framework is presented and detail in this section.

A. Notation

Considering the fact that the proposal framework is applied to multiple robots n with different coordinates frames and during the formulation we will require coordinate transformation matrix from one coordinate (designated by *from*) to another coordinate frame (designated by *to*), we use the following notation: $^{to}_{from}\mathrm{S}_\mathrm{n}$.To represent the coordinate transformation, we label $\{C\}$ for camera frame, $\{B\}$ for body frame, $\{N\}$ for navigation expressed in *ENU* (earth-fixed east-north-up) and $\{W\}$ for global frame expressed in *ECEF* (earth-centered, earth-fixed) coordinate. The upper bold case notation represent matrix, lower bold case vectors and lower case scalar variables.

B. System Overview

The proposed cooperative stereo framework architecture is outlined in figure 2.

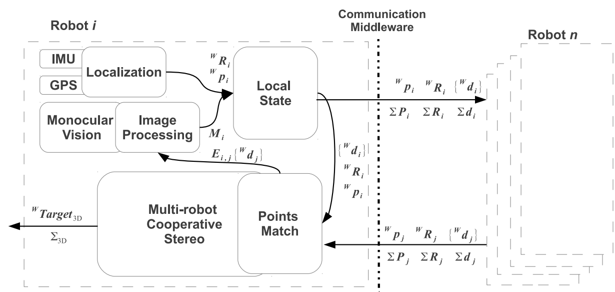

Fig. 2. Multi-Robot Cooperative Stereo Architecture

The architecture is composed by a localization layer responsible for providing to the local state layer the attitude and global frame position of the robot. Although in the current implementation this information came from INS/GPS fusion but could be in the future come from any other localization system. For each camera available in the robot, a image processing block provide the $\{M_i\}$ with the detected target measurement and the correspond uncertainty. For each image processed the robot share over a middleware communication, with the robot that is sharing the same overlap view, the position $^W\mathbf{p}_\mathrm{i}$ and orientation of camera $^W_C\mathbf{R}_\mathrm{i}$ as well as a list of possible targets measurements $\{^W\mathbf{d}_\mathrm{i}\}$. The information provided from other robots is then used by a features correspond block (section II-C). The position and attitude of both robots cameras will estabilish the essential matrix $\mathbf{E}_{i,j}$ that will define epipolar restrictions between targets pairs. Finally the target pairs are applied to obtain the 3D target measurement $^W\mathbf{Target}_{\mathrm{3D}}$ and uncertainty Σ_{3D} as detailed in algorithm 1. Before present the proposed algorithm we will describe for robot i the variables and inputs depicted in figure 3.

The camera position in the global frame is obtained:

$$^W\mathbf{p}_\mathrm{i} = ^W_N\mathbf{T}_\mathrm{i}.^N_B\mathbf{T}_\mathrm{i}.^B\mathbf{p}_\mathrm{i} \tag{1}$$

where $^B\mathbf{p}_\mathrm{i}$ is the camera position in body frame as:

$$^B\mathbf{p}_\mathrm{i} = ^B_C\mathbf{R}_\mathrm{i}.(-\mathbf{t}_i) \tag{2}$$

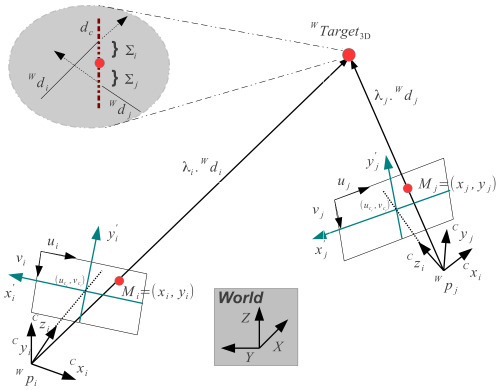

Fig. 3. System setup between robot i and j

being ${}^{B}_{C}\mathbf{R}_{\rm i}$ and $\mathbf{t}_i$ achieved from the camera extrinsic calibration and where the ${}^{W}_{N}\mathbf{T}_{\rm i} = \begin{bmatrix} {}^{N}_{B}\mathbf{R} & {}^{W}_{robot}\mathbf{P} \end{bmatrix}$ and ${}^{N}_{B}\mathbf{T}_{\rm i} = \begin{bmatrix} {}^{N}_{B}\mathbf{R} & 0 \end{bmatrix}$ are respectively the transformation matrix from body to navigation and navigation to world.

The direction vector to the target in world frame is:

$$ {}^{W}\mathbf{d}_{\rm i} = {}^{W}_{N}\mathbf{R}_{\rm i}.{}^{N}_{B}\mathbf{R}_{\rm i}.{}^{B}\mathbf{d}_{\rm i} \tag{3} $$

where ${}^{B}\mathbf{d}_{\rm i}$ is the same vector in body frame equal to ${}^{B}_{C}\mathbf{R}_{\rm i}.[\frac{x_i - \mathrm{u_c}}{f_x}, \frac{y_i - \mathrm{v_c}}{f_y}, 1]^T$, (u_c, v_c) principal point (that is usually at the image center), focal lengths (f_x, f_y) and $\{M_i\} = (x_i, y_i)$ are the detected target measurement.

C. Multi-Robot Stereo Correspondence

The features match between different cameras i, j is performed though the epipolar geometric line information. To avoid ambiguous matches, the corresponding points are searched over the epipolar line in a narrow band within 2σ distance. In order to obtain the epipolar line, each robot will share the rotation matrix (4) and the candidate feature position (1), both related to the global frame.

$$ {}^{W}_{C}\mathbf{R}_{\rm i} = {}^{W}_{N}\mathbf{R}_{\rm i}.{}^{N}_{B}\mathbf{R}_{\rm i}.{}^{B}_{C}\mathbf{R}_{\rm i} \tag{4} $$

With this information we estimate the stereo rotation R matrix and translation t vector 5 and consequently the essencial matrix $E_{i,j} = \hat{t}.R$.

$$ \begin{cases} t = ({}^{W}_{C}\mathbf{R}\mathrm{i})^T.({}^{W}\mathbf{p}_{\rm i} - {}^{W}\mathbf{p}_{\rm j}) \\ \\ \mathbf{R} = ({}^{W}_{C}\mathbf{R}_{\rm j})^T.{}^{W}_{C}\mathbf{R}_{\rm i} \end{cases} \tag{5} $$

D. Stereo Measurement Uncertainty

In order to define uncertainty in 3D target we will first define the uncertainty associated to each intersection point P_{int_i} and P_{int_j}. To achieve the P_{int_i} covariance called $\Sigma_{P_{int_i}}$, we need to obtain the jacobian matrix of P_{int_i} in order to input variables $\nu_{i,j}$.

$$ \begin{aligned} J_i &= \nabla_{\nu_{i,j}} P_{int_i}(\nu_{i,j}) \\ J_j &= \nabla_{\nu_{i,j}} P_{int_j}(\nu_{i,j}) \end{aligned} \tag{6} $$

Algorithm 1 Multi-Robot Cooperative Stereo

Assuming that each robot share a 3-tuple $({}^{W}\mathbf{p}_{\rm i}, {}^{W}_{C}\mathbf{R}_{\rm i}, \{{}^{W}\mathbf{d}_{\rm i}\})$ for robot i and $({}^{W}\mathbf{p}_{\rm j}, {}^{W}_{C}\mathbf{R}_{\rm j}, \{{}^{W}\mathbf{d}_{\rm j}\})$ for robot j. For each pair of points received from robots i, j we will perform the following steps:

Step 1: Evaluate the correspondence between points detected in each camera considering the epipolar constraint (details in section II-C). If the points are without correspond with the epipolar constraint, the algorithm proceed to the next steps otherwise the tuples are label as being targets.

Step 2: Obtain perpendicular vector to ${}^{W}\mathbf{d}_{\rm i}$ and ${}^{W}\mathbf{d}_{\rm j}$ (see figure 4).

$\mathbf{d}_c = \perp ({}^{W}\mathbf{d}_{\rm i}, {}^{W}\mathbf{d}_{\rm j})$

Step 3: Estimate the value of the λ_i where the ray $({}^{W}\mathbf{p}_{\rm i} + \lambda_i.{}^{W}\mathbf{d}_{\rm i})$ intersects the plane π_j defined by the other monocular robot camera j optical center in world frame ${}^{W}\mathbf{p}_{\rm j}$ and the direction vector $({}^{W}\mathbf{d}_{\rm j}, \mathbf{d}_c)$ being the intersection point $\mathbf{P}_{int_i} = {}^{W}\mathbf{p}_{\rm j} + \lambda_j.{}^{W}\mathbf{d}_{\rm j}$. The same approach for λ_j.

$$ \begin{cases} \lambda_i = \frac{({}^{W}\mathbf{p}_{\rm j} - {}^{W}\mathbf{p}_{\rm i})^T.(\mathbf{d}_c \wedge {}^{W}\mathbf{d}_{\rm i})}{{}^{W}\mathbf{d}_{\rm i}{}^T.(\mathbf{d}_c \wedge {}^{W}\mathbf{d}_{\rm i})} \\ \\ \lambda_j = \frac{({}^{W}\mathbf{p}_{\rm i} - {}^{W}\mathbf{p}_{\rm i})^T.(\mathbf{d}_c \wedge {}^{W}\mathbf{d}_{\rm j})}{{}^{W}\mathbf{d}_{\rm j}{}^T.(\mathbf{d}_c \wedge {}^{W}\mathbf{d}_{\rm j})} \end{cases} $$

Step 4: Obtain 3D target point in ECEF coordinate frame (section II-D)

$$ {}^{W}\mathbf{Target}_{\rm 3D} = \frac{\Sigma_{Pint_j}}{\Sigma_{Pint_i} + \Sigma_{Pint_j}}.({}^{W}\mathbf{p}_{\rm i} + \lambda_i.{}^{W}\mathbf{d}_{\rm i}) + \frac{\Sigma_{Pint_i}}{\Sigma_{Pint_i} + \Sigma_{Pint_j}}.({}^{W}\mathbf{p}_{\rm j} + \lambda_j.{}^{W}\mathbf{d}_{\rm j}) $$

Step 5: Evaluation the Euclidean distance between two points projected in the global frame in case of λ_i and λ_j are positive . The thr value is a metric physical distance in mm.

if $||({}^{W}\mathbf{p}_{\rm j} + \lambda_j.{}^{W}\mathbf{d}_{\rm j}) - ({}^{W}\mathbf{p}_{\rm i} + \lambda_i.{}^{W}\mathbf{d}_{\rm i})|| < thr$ **then**
  return ${}^{W}\mathbf{Target}_{\rm 3D}$
end if

where the input state vector is defined as:

$$ \nu_{i,j} = \left[\overbrace{P_i, \mathbf{R}_i, d_i}^{Robot_i}, \overbrace{P_j, \mathbf{R}_j, d_j}^{Robot_j} \right] \tag{7} $$

With the jacobian we can combine the uncertainty in the state variables $N_{i,j}$ to covariance of P_{int_i} and P_{int_j}.

$$ \mathbf{N}_{i,j} = diag[\Sigma_{P_i}, \Sigma_{R_i}, \Sigma d_i, \Sigma_{P_j}, \Sigma_{R_j}, \Sigma d_j] \tag{8} $$

$$ \begin{aligned} \Sigma_{P_{int_i}} &= J_i.\mathbf{N}_{i,j}.J_i^T \\ \Sigma_{P_{int_j}} &= J_j.\mathbf{N}_{i,j}.J_j^T \end{aligned} \tag{9} $$

As seen in the algorithm the covariance of the both intersection points is used in the determination of the 3D measurement by weighting the uncertainty of each of them (see figure 4) in opposition to classic stereo mid-point triangulation method.

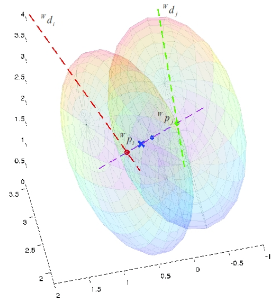

Fig. 4. Snapshot from the covariance 3D ellipse of P_{int_i} and P_{int_j} and the intersection point obtained from the algorithm. Mid-Point Triangulation method (blue dot). Triangulation based on the state covariance value (blue cross). Perpendicular vector ${}^{W}\mathbf{d}_c$ to $({}^{W}\mathbf{d}_i, {}^{W}\mathbf{d}_j)$ (purple line).

III. Experimental Setup

A. Outdoor Scenario

To evaluate the proposed multi-robot collaborative stereo the chosen experimental scenario was a non-urban area with several landscape elements, e.g., vegetation, water, rocks, bushes and some semi-urban structures such as gravel paths.

Fig. 5. Experimental Scenario

Fig. 6. Static target tracking by the robot TIGRE

The target tracking used during the experimental tests was an orange life jacket (see figure 6) with a size of $37cm \times 67cm$ equipped with a RTK GPS Septentrio L1 L2 able to provide in post-process a centimeter-accuracy lower than $10cm$. This will allow to evaluate the results from the cooperative stereo and consider the target position as a external ground-truth.

B. Vehicles

The robot TIGRE (see figure 7) is an autonomous ground robot for exploration and activity in unstructured environments. The vehicle has electric propulsion and is equipped with an on board processing Quad Core Intel(R) Core(TM) i5 CPU 750 @ 2.67GHz, 4GB RAM, running a Linux operating system, wireless communications, infra-red thermographic camera, laser rangefinder, two visible spectrum cameras in a rigid stereo baseline (~ 0.76 meters) with a pixel resolution of 1278×958, Novatel GPS receiver and IMU Microstrain.

The MAV (Micro Aerial Vehicle) (see figure 8) is a helicopter driven by four rotors, symmetric to the center of

Fig. 7. TIGRE - Terrestrial Intelligent Ground Robotic Explorer

Fig. 8. Asctec@ Pelican MAV

mass equipped with a Flight Control Unit (FCU) for data fusion (GPS and IMU) and flight control, an onboard 1.6 GHz Intel Atom Based Embedded Computer, 802.11n Wifi and a monocular camera from IDS UEYe LE with a resolution of 1280×1024. Both vehicles are running Linux and the ROS framework as a middleware for communication, parameters and monitoring of all processes. It is also crucial for the whole system to work the accurate time synchronization between all robots involved in the cooperative stereo.

IV. Results

In this section we describe the results obtained from two experimental cases that were performed in an outdoor scenario with a static target. The fact that we are using a static target was due to the importance of evaluating the quality of results from stereo triangulation with a rigid baseline (IV-A) and the paper proposal method with a multi-robot collaborative stereo (IV-B) in a similar context able to be reproduced.

A. Experiment I: Stereo rigid known baseline

For this experimental case, a stereo rigid baseline available at TIGRE was used to track the target. This means that the MAV was not available, so the results will express the quality of perception from TIGRE that was at a distance of ~ 35 meters from the static target (see figure 6) and moving towards with speed of $0.4\ m/s$.

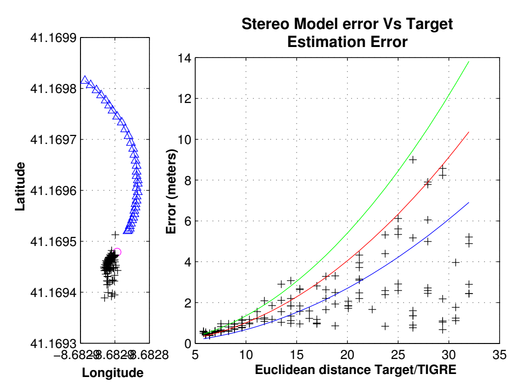

Fig. 9. Target estimation error for a stereo rigid known baseline. **Left:** TIGRE GPS trajectory (blue triangle), target GPS RTK position (magenta circle). Estimate position of the target (black cross). **Right:** Estimation position error related to the target (black cross) compared with the stereo model error (green, red and blue lines).

As expected and considering the reference[15], the perception accuracy of the TIGRE target tracking followed the stereo model error: $\epsilon_z = \frac{z^2}{b.f}.\epsilon_d$ where ϵ_z is the depth error, z is the depth, b the baseline, f the focal length in pixels and ϵ_d the matching error in pixels. The stereo model error is expressed in the lines from figure 9 on the right to different values of ϵ_d and the black crosses the estimation position error related to the target. Figure 10 presents the stereo vision covariance for three instances related to the target position. We observe that the covariance decreases with shortening the distance and the bearing angle is consistent even for large distances. It became clear that was not possible to have a good accuracy for target tracking with local perception due to the normally ($\sim$ 1 m) available rig baseline.

B. Experiment II: Multi-Robot Cooperative Stereo

Supported by the monocular MAV camera both robots are able to obtain a flexible stereo baseline using the proposed multi-robot collaborative stereo framework detail in section II. The experiment was composed of several steps: TIGRE detected the target and shared the estimation position to MAV, MAV moved based on the information provided by TIGRE to the top of the target and remained on the top based on local perception, MAV shared a 3-tuple $({}^W\mathbf{p}_{\mathrm{i}}, {}^W_C\mathbf{R}_{\mathrm{i}}, \{{}^W\mathbf{d}_{\mathrm{i}}\})$ to TIGRE in order to in a cooperative way estimate 3D target position. Results are showed in figures 11 and 12.

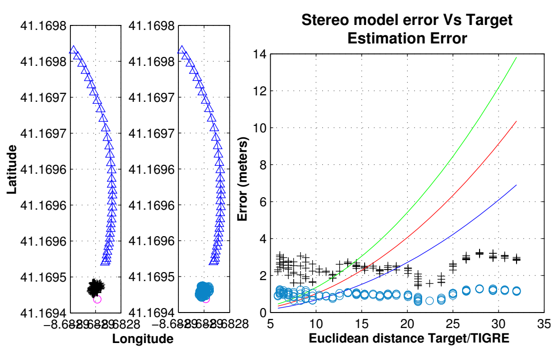

Fig. 11. Target estimation error with Multi-Robot Cooperative Stereo. **Left:** TIGRE GPS trajectory (blue triangle), target GPS RTK position (magenta circle). Estimate position of the target for each method: Mid-Point Triangulation (black cross) and State Covariance Sigma Value (blue circle). **Right:** Estimation position error related to the target with Mid-Point Triangulation Method (black cross) and State Covariance Sigma Value Method (blue circle) compared with the stereo model error (green, red and blue lines).

Comparing the results from section IV-A with IV-B, it is possible to observe that the proposal framework with cooperative dynamic stereo baseline reduced dramatically the target estimation error. This improvement is even more noticeable if we compare the results between figures 9 and 11 when applied the method based on state covariance sigma value (blue circle) detailed in section II-D. From figure 12 we can observe that the resulting uncertainty in 3D target is dominated by the MAV uncertainty mainly caused by the low cost GPS error ($\sim$ 2 m).

V. Conclusions

In this work we present a framework for multi-robot realtime 3D high dynamic target estimation in outdoor scenarios. The proposed framework provides the following functionalities:

- Determination of 3D target measurement and associated uncertainty from image measurements from two cameras in robots and robots localization as well as the associated uncertainties;
- Mechanism to help the target search and identification in the image processing blocks in robots with low computational capacities;
- Mechanism to help the matching and association of 2D targets
- Better understanding of how the several sources of uncertainty contributes to measurement uncertainty

We demonstrate the advantages of this approach by comparing a stereo rigid baseline standalone target tracking with the proposed multi-robot cooperative stereo between a micro aerial vehicle (MAV) and an autonomous ground vehicle (AGV). The Field experimental cases, show that our proposal framework with cooperative dynamic stereo baseline reduces dramatically the target estimation error. This novelty will allow in future to establish an information framework for the formation control of multi-robot system. Additionally, this cooperative perception framework when integrated in a multi target tracking architecture, like a DDF, will endows it with a fast track initialization and more robust data association layer in highly dynamic scenarios.

Acknowledgments

The authors acknowledge the support given by the ISEP-IPP and INESC TEC to this project. This work is financed by the ERDF - European Regional Development Fund through the COMPETE Programme and by National Funds through the FCT - Portuguese Foundation for Science and Technology within project FCOMP-01-0124-FEDER-022701 and under Perception-Driven Coordinated Multi-Robot Motion Control Project with reference PTDC/EEA-CRO/100692/2008.

References

[1] Markus W. Achtelik, Stephan Weiss, Margarita Chli, Frank Dellaerty, and Roland Siegwart. Collaborative stereo. In *Proc. IEEE/RSJ Int Intelligent Robots and Systems (IROS) Conf*, pages 2242–2248, 2011.

[2] D.Blake Barber, JoshuaD. Redding, TimothyW. McLain, RandalW. Beard, and ClarkN. Taylor. Vision-based target geo-location using a fixed-wing miniature air vehicle. *Journal of Intelligent and Robotic Systems*, 47, 2006.

[3] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29 (6):1052 – 1067, June 2007.

[4] G. Di Leo, C. Liguori, and A. Paolillo. Covariance propagation for the uncertainty estimation in stereo vision. *Instrumentation and Measurement, IEEE Transactions on*, 60(5):1664 – 1673, May 2011.

[5] Brian J. Julian, Michael Angermann, Mac Schwager, and Daniela Rus. A scalable information theoretic approach to distributed robot coordination. In *Intelligent Robots*

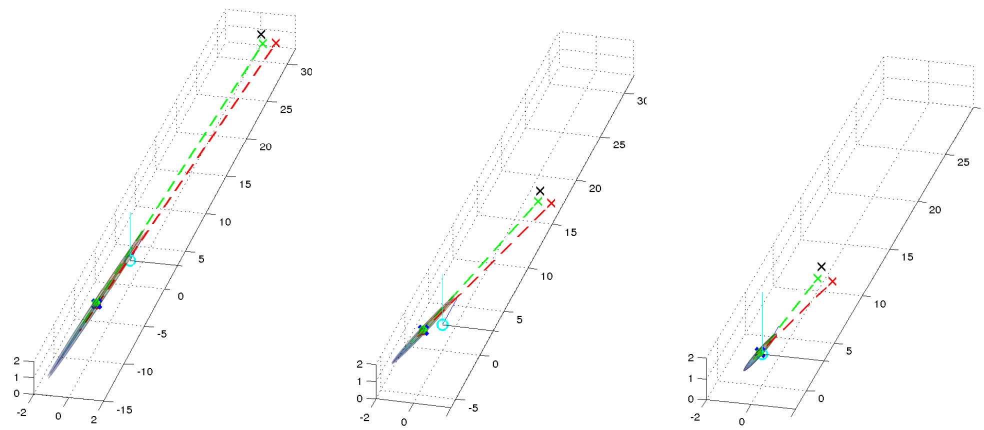

Fig. 10. Snapshot from the stereo vision covariance 3D ellipse for three instances of the experiment I. Red and green crosses are respectively the left and right camera position related to the global frame. The blue circle represent the TARGET RTK GPS position.

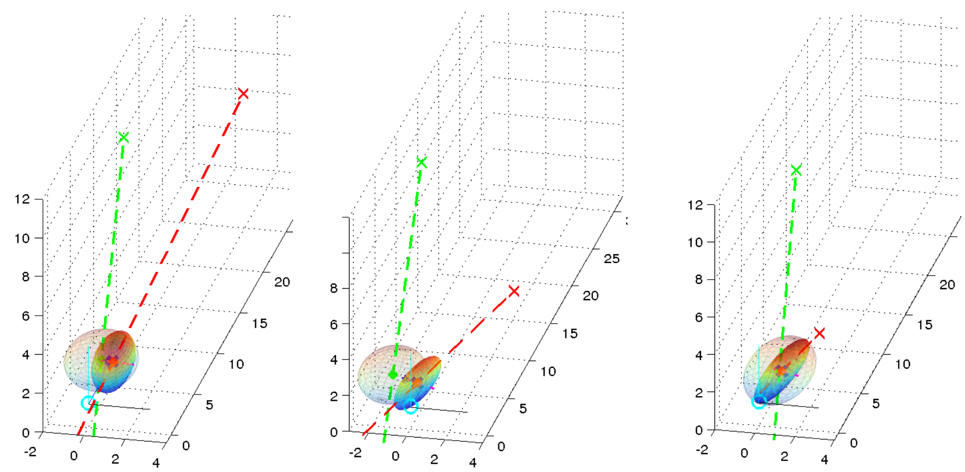

Fig. 12. Snapshot from the stereo vision covariance 3D ellipse for three instances of the experiment II. Red and green crosses are respectively the left and right camera position related to the global frame. The blue circle represent the TARGET RTK GPS position.

and Systems (IROS), 2011 IEEE/RSJ International Conference on, pages 5187 –5194, Sept 2011.

[6] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234, 2007.

[7] Aleksandr Kushleyev, Vijay Kumar, and Daniel Mellinger. Towards a swarm of agile micro quadrotors. In *Proceedings of Robotics: Science and Systems*, Sydney, Australia, July 2012.

[8] A. Marino, F. Caccavale, L.E. Parker, and G. Antonelli. Fuzzy behavioral control for multi-robot border patrol. In *Control and Automation, 2009. MED '09. 17th Mediterranean Conference on*, pages 246 – 251, 2009.

[9] L.-L. Ong, T. Bailey, H. Durrant-Whyte, and B. Upcroft. Decentralised particle filtering for multiple target tracking in wireless sensor networks. In *Information Fusion 11th International Conference on*, pages 1 – 8, July 2008.

[10] M. Schwager, B. J. Julian, M. Angermann, and D. Rus. Eyes in the sky: Decentralized control for the deployment of robotic camera networks. 99(9):1541–1561, 2011.

[11] J.H. Sutcliffe, P. Riseborough, and H. Durrant-Whyte. Decentralised data fusion applied to a network of unmanned aerial vehicles. In *Information, Decision and Control*, 2002.

[12] Fumin Zhang and N.E. Leonard. Cooperative filters and control for cooperative exploration. *Automatic Control, IEEE Transactions on*, 55(3):650 – 663, March 2010.

[13] Ke Zhou and S.I. Roumeliotis. Multirobot active target tracking with combinations of relative observations. *Robotics, IEEE Transactions on*, 27(4):678 – 695, 2011.

[14] Danping Zou and Ping Tan. Coslam: Collaborative visual slam in dynamic environments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(2):354 – 366, Feb. 2013.

[15] Frahm J.M. Mordohai P. Pollefeys M. ï¿£Gallup, D. Variable baseline/resolution stereo. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.