2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Passivity-based Cooperative Estimation of 3D Target Motion for Visual Sensor Networks: Analysis on Averaging Performance

Takeshi Hatanaka and Masayuki Fujita

***Abstract*— This paper presents a novel cooperative estimation algorithm for visual sensor networks. We consider the situation where multiple smart vision cameras with computation and communication capability see a group of target objects. The objective of the present algorithm is to meet two requirements: averaging and tracking. In order to meet the requirements simultaneously, we present a cooperative estimation algorithm based on passivity of the kinematic model of rigid body motion. We then provide an upper bound of the ultimate error between the actual average and the estimates given by the present algorithm. Finally the effectiveness of the present estimation algorithm is demonstrated through experiments.**

I. Introduction

A visual sensor network [1], [2] is a network consisting of spatially distributed smart cameras, which is a kind of sensor networks. Unlike the other sensors measuring some values such as temperature and pressure, vision sensors do not provide explicit data but combining image processing techniques or human operators gives rich information which the other sensors do not provide. Due to the nature, visual sensor networks are useful in environmental monitoring, surveillance, target tracking and entertainment.

A lot of research works have been devoted to combining control techniques with visual information so-called visual feedback control or images in the loop [3]–[7]. Among them, we focus on an estimation problem of 3D rigid body motion. In visual sensor networks, it is expected not only to give an estimate but also to cooperate with each other camera in a distributed fashion, which brings us new theoretical challenges. The main advantages of cooperation are: (i) accuracy of estimation by integrating rich information, (ii) tolerance against obstruction, misdetection in image processing and sensor failures and (iii) wide vision and elimination of blind areas by fusing images of a scene from a variety of viewpoints. To tackle such distributed control and estimation problems, cooperative control and estimation schemes as in [8]–[12] provide useful methodologies.

[12] presents distributed Kalman filters for sensor networks based on the consensus algorithm [8]. Unfortunately, the algorithm is not applicable to our problem since the object's pose takes values in a non-Euclidean space. Meanwhile, [13], [14] present distributed estimation algorithms for visual sensor networks, where [13] computes so-called Riemanian mean and [14] Euclidean mean which are averages on Special Orthogonal Group [15]. However, [13] focuses on the averaging by assuming that the target's orientation is obtained *a priori* and does not mention estimation from vision data. Though [14] considers both estimation and averaging and provides an upper bound on the error between the actual mean and the estimates, the bound cannot be computed *a priori* and can be obtained only after gaining estimates.

Takeshi Hatanaka and Masayuki Fujita are with the Department of Mechanical and Control Engineering, Tokyo Institute of Technology, Tokyo 152-8552, JAPAN `hatanaka@ctrl.titech.ac.jp`

In this paper, we present a novel cooperative estimation algorithm for visual sensor networks. We consider the situation where multiple vision cameras with computation and communication capability see a group of target objects. The objective of the present algorithm is to meet two requirements: (i) gaining estimates close to an average pose for static objects and (ii) tracking of the estimates to moving objects' poses. For this purpose, we first present a cooperative estimation algorithm based on the passivity-based visual motion observer [7] and passivity-based pose synchronization law [10]. Then, we provide an upper bound of the ultimate error between the actual average and the estimates given by the present algorithm. The result gives an insight into the relation between the mean estimation accuracy and the feedback gain in the visual motion observer, namely the estimate becomes accurate as the gain becomes small. The conclusion basically corresponds to that in works on multi-agent optimization [16] if the estimation problem is viewed as a distributed optimization problem on $SE(3)$. However, we see that unlike optimization problems on a vector space an offset occurs with respect to the orientation estimates in $SO(3)$ regardless of the gain selection. We finally show the effectiveness of the present algorithm through experiments.

We finally give some notations used in this paper, where the readers are recommended to refer to [17] for details on the terminologies. Throughout this paper, we use the notation $e^{\hat{\xi}\theta_{ab}} \in \mathcal{R}^{3\times 3}$ to represent the rotation matrix of a frame Σ_b relative to a frame Σ_a, which is orthogonal with unit determinant and hence an element of the Lie group $SO(3)$. $\xi_{ab} \in \mathcal{R}^3$ specifies the rotation axis and $\theta_{ab} \in \mathcal{R}$ is the rotation angle. For simplicity we use $\xi\theta_{ab}$ to denote $\xi_{ab}\theta_{ab}$. The notation '$\wedge$' is the operator such that $\hat{a}b = a \times b$ for the vector cross-product $\times$. The vector space of all 3×3 skew-symmetric matrices is denoted $so(3)$. The notation '$\vee$' denotes the inverse operator to '$\wedge$'. We use $g_{ab} = \begin{bmatrix} e^{\hat{\xi}\theta_{ab}} & p_{ab} \\ 0 & 1 \end{bmatrix}$ as the homogeneous representation of $g_{ab} = (p_{ab}, e^{\hat{\xi}\theta_{ab}}) \in SE(3) := \mathcal{R}^3 \times SO(3)$ describing the configuration of Σ_b relative to Σ_a. The adjoint transformation associated with g_{ab} is denoted by $\mathrm{Ad}_{(g_{ab})}$. Similarly to the

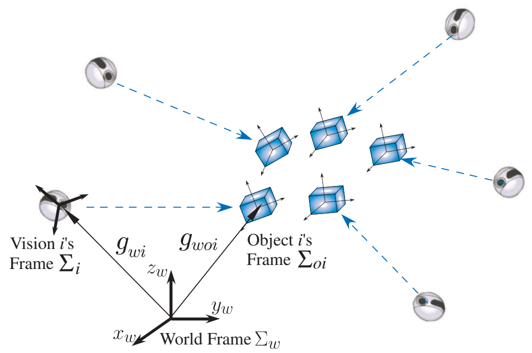

Fig. 1. Visual Sensor Networks

definition of $so(3)$, we define $se(3) := \{(v, \hat{\omega}) : v \in \mathcal{R}^3, \hat{\omega} \in so(3)\}$. In homogeneous representation, we write an element $V := (v, \omega)$ of $se(3)$ as $\hat{V} = \begin{bmatrix} \hat{\omega} & v \\ 0 & 0 \end{bmatrix}$.

II. Visual Sensor Networks

A. Relative Rigid Body Motion

Throughout this paper, we consider the situation where n vision cameras see a group of target objects (Fig. 1). The motivation to consider the situation will be explained in Sections II-D. Suppose that each vision camera $i \in \mathcal{V} := \{1, \cdots, n\}$ has communication and computation capability.

Let the coordinate frames Σ_w, Σ_i and Σ_{o_i} represent the world frame, the i-th vision camera frame, and the frame of the object which vision camera i sees, respectively. Then, the pose of vision camera i and object o_i relative to Σ_w are denoted by $g_{wi} = (p_{wi}, e^{\hat{\xi}\theta_{wi}})$ and $g_{wo_i} = (p_{wo_i}, e^{\hat{\xi}\theta_{wo_i}})$. Let $p_{io_i} \in \mathcal{R}^3$ and $e^{\hat{\xi}\theta_{io_i}} \in SO(3)$ be the position vector and the rotation matrix from the vision camera frame Σ_i to the object frame Σ_{o_i}. Then, the pose of Σ_{o_i} relative to Σ_i can be represented by $g_{io_i} = (p_{io_i}, e^{\hat{\xi}\theta_{io_i}}) \in SE(3)$ and satisfies $g_{io_i} = g_{wi}^{-1} g_{wo_i}$.

We next define the body velocity of the object Σ_{o_i} relative to the world frame Σ_w as $V^b_{wo_i} = (v_{wo_i}, \omega_{wo_i})$ or

$$\hat{V}^b_{wo_i} = g^{-1}_{wo_i} \dot{g}_{wo_i} = \begin{bmatrix} \hat{\omega}_{wo_i} & v_{wo_i} \\ 0 & 0 \end{bmatrix} \in \mathcal{R}^{4\times 4}, \quad (1)$$

where v_{wo_i} and ω_{wo_i} represent the linear velocity of the origin and the angular velocity of Σ_{o_i} relative to Σ_w, respectively [17]. Similarly, vision camera i's body velocity relative to Σ_w will be denoted as $V^b_{wi} = (v_{wi}, \omega_{wi})$.

By using the body velocities V^b_{wi} and $V^b_{wo_i}$, the body velocity of the relative rigid body motion g_{io_i} is written as

$$V^b_{io_i} = -\mathrm{Ad}_{(g^{-1}_{io_i})} V^b_{wi} + V^b_{wo_i}. \quad (2)$$

Equation (2) is a standard formula for the relation between the body velocities of three coordinate frames [17].

B. Visual Measurement

In this subsection, we define the visual measurement of the vision camera which is available for estimation of target objects' motion. Throughout this paper, we use the pinhole camera model with a perspective projection [17].

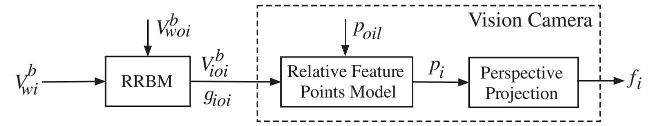

Fig. 2. Block Diagram of the RRBM with Vision Camera (RRBM is an acronym for Relative Rigid Body Motion)

In this paper, we assume that each target object has m feature points and each vision camera can extract them from the visual data by using some techniques. The position vectors of the target object i's l-th feature point relative to Σ_{o_i} and Σ_i are denoted by $p_{o_il} \in \mathcal{R}^3$ and $p_{il} \in \mathcal{R}^3$ respectively. Using a transformation of the coordinates, we have $p_{il} = g_{io_i} p_{o_il}$, where p_{o_il} and p_{il} should be regarded, with a slight abuse of notation, as $[p^T_{il}\ 1]^T$ and $[p^T_{o_il}\ 1]^T$.

Let the m feature points of the object o_i on the image plane coordinate, denoted by $f_i := [f^T_{i1} \ \cdots \ f^T_{im}]^T \in \mathcal{R}^{2m}$, be the measurement of the camera i. It is well known [17] that the perspective projection of the l-th feature point onto the image plane gives us the image data $f_{il} \in \mathcal{R}^2$ as

$$f_{il} = \lambda_i \begin{bmatrix} x_{il}/z_{il} & y_{il}/z_{il} \end{bmatrix}, \quad (3)$$

where $p_{il} = [x_{il}\ y_{il}\ z_{il}]^T$ and λ_i is a focal length of vision camera i. It is straightforward to extend this model to m image points f_i and $p_i := [p^T_{i1} \ \cdots \ p^T_{im}]^T \in \mathcal{R}^{3m}$. Under the assumption that each vision camera i knows $p_{o_il} \in \mathcal{R}^3$, the visual measurement f_i depends only on the relative rigid body motion g_{io_i}. Fig. 2 shows the block diagram of the relative rigid body motion with the camera model.

C. Communication Model

The vision cameras have communication capability with the neighboring cameras and constitute a network. The communication is modeled by a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$. Namely, vision camera i can get some information from j if $(i, j) \in \mathcal{E}$. In addition, we define the neighbor set $\mathcal{N}_i := \{j \in \mathcal{V} |\ (i, j) \in \mathcal{E}\}$. Now, we assume the following.

Assumption 1: The communication graph G is fixed, balanced and strongly connected.

D. Objectives

The main objective of this paper is to present a cooperative estimation algorithm achieving

- Averaging: Each vision camera i estimates a pose close to an average of $\{g_{io_j}\}_{j\in\mathcal{V}}$ for static objects.
- Tracking: The estimates track the pose g_{io_i} for moving objects with a finite tracking error.

Though the present algorithm embodies the tracking nature from its structure, our theoretical interests are restricted to the averaging in this paper.

The problem is motivated by scenarios such as estimation of group behaviors, estimation under uncertainties including noises, incomplete localization and parametric uncertainties of vision cameras. For example, suppose that multiple vision cameras see a single target object, where each camera has uncertainties. In theory, the feature points on the image plane are given by Equation (3). However, as exemplified in experiments [14], it is not always true due to parametric uncertainties of the cameras and distortion of lenses. In such a situation, all the individual cameras can do are to estimate a pose consistent with the visual measurement. Since all the cameras have uncertainties, the situation is interpreted as if vision cameras see different target objects. A way to gain an accurate estimate under the situation is averaging the uncertain information among cameras [12].

Let us finally introduce the following means with respect to positions and orientations as an average of $\{g_{wo_j}\}_{j\in\mathcal{V}}$. We employ the arithmetic mean $p^* = \frac{1}{n}\sum_{j\in\mathcal{V}} p_{wo_j}$ as the position average of $\{p_{wo_j}\}_{j\in\mathcal{V}}$, and a so-called Euclidean mean [15] as the orientation average of $\{e^{\hat{\xi}\theta_{wo_j}}\}_{j\in\mathcal{V}}$. In order to define the Euclidean mean, we introduce the energy function on $SO(3)$ $\phi(e^{\hat{\xi}\theta}) = \mathrm{tr}(I_3 - e^{\hat{\xi}\theta}) = \frac{1}{2}\|I_3 - e^{\hat{\xi}\theta}\|_F^2$, where $\|M\|_F$ is the matrix Frobenius norm of matrix M. Then, the Euclidean mean of $\{e^{\hat{\xi}\theta_{wo_j}}\}_{j\in\mathcal{V}}$ is defined by

$$e^{\hat{\xi}\theta^*} := \arg\min_{e^{\hat{\xi}\theta}\in SO(3)} \sum_{j\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta} e^{\hat{\xi}\theta_{wo_j}}). \tag{4}$$

Hereafter, we use the notations $g^* = (p^*, e^{\hat{\xi}\theta^*})$ and $g_i^* = (p_i^*, e^{\hat{\xi}\theta_i^*}) := g_{wi}^{-1} g^*$.

III. Visual Motion Observer

In this section, we consider the problem that a vision camera i estimates the target object motion g_{io_i} from the visual measurements f_i without considering communication. For this purpose, we introduce the visual motion observer presented in [7].

We first prepare a model of the actual rigid body motion (2) similarly to the Luenberger-type observer. as

$$\bar{V}_{io_i}^b \;\; = \;\; -\mathrm{Ad}_{(\bar{g}_{io_i}^{-1})} V_{wi}^b + u_{ei}, \tag{5}$$

where $\bar{g}_{io_i} = (\bar{p}_{io_i}, e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})$ and $\bar{V}_{io_i}^b$ are estimates of the relative pose g_{io_i} and its velocity respectively The input $u_{ei} = [v_{uei}^T \ \omega_{uei}^T]^T$ is to be determined to drive the estimated values $\bar{g}_{io_i}$ and $\bar{V}_{io_i}^b$ to their actual values. Once the estimate $\bar{g}_{io_i}$ is determined, the estimated measurement $\bar{f}_i$ is also computed by (3).

In order to establish the estimation error system, we define the estimation error between the estimated value $\bar{g}_{io_i}$ and the actual relative rigid body motion g_{io_i} as $g_{ei} = (p_{ei}, e^{\hat{\xi}\theta_{ei}}) := \bar{g}_{io_i}^{-1} g_{io_i}$. Using the notations $e_R(e^{\hat{\xi}\theta}) := \mathrm{sk}(e^{\hat{\xi}\theta})^\vee$ and $\mathrm{sk}(e^{\hat{\xi}\theta}) := \frac{1}{2}(e^{\hat{\xi}\theta} - e^{-\hat{\xi}\theta})$, the vector representation of the estimation error is given by

$$e_{ei} := E_R(g_{ei}), \ E_R(g_{ei}) := \left[\begin{array}{cc} p_{ei}^T & e_R^T(e^{\hat{\xi}\theta_{ei}}) \end{array}\right]^T. \tag{6}$$

Note that $e_{ei} = 0$ iff $g_{io_i} = \bar{g}_{io_i}$.

If we define the visual measurement error as $f_{ei} := f_i(g_{io_i}) - \bar{f}_i(\bar{g}_{io_i})$, then the relation between the actual vision data and the estimated one can be approximately given by $f_{ei} = J_i(\bar{g}_{io_i}) e_{ei}$ [7], where $J_i(\bar{g}_{io_i}) : SE(3) \to \mathcal{R}^{2m\times 6}$ is the well-known image Jacobian. Now, if $m \geq 4$, the estimation error vector is reconstructed by

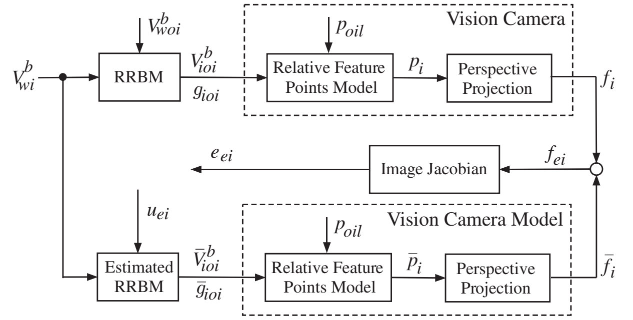

Fig. 3. Estimation Error System

$$e_{ei} = J_i^\dagger(\bar{g}_{io_i}) f_{ei}, \tag{7}$$

where $\dagger$ denotes the pseudo-inverse.

Differentiating $g_{ei} = \bar{g}_{io_i}^{-1} g_{io_i}$ with respect to time and using (2) and (5), we obtain the estimation error system

$$V_{ei}^b \;\; = \;\; -\mathrm{Ad}_{(g_{ei}^{-1})} u_{ei} + V_{wo_i}^b. \tag{8}$$

Fig. 3 shows the block diagram of the system (8).

We next form the visual motion observer and analyze stability of the closed-loop system based on passivity. [7] proves that if $V_{wo_i}^b = 0$, then the estimation error system (8) is passive from the input u_{ei} to the output $-e_{ei}$. Based on the passivity property, we consider the following input

$$u_{ei} = -k_{ei}(-e_{ei}) = k_{ei} e_{ei}, \ k_{ei} > 0. \tag{9}$$

Then, we have the following facts from passivity-based control theory.

Fact 1: [7] (i) If $V_{wo_i}^b = 0$, then the equilibrium point $e_{ei} = 0$ for the closed-loop system (8) and (9) is asymptotically stable. (ii) Given a positive scalar $\tilde{\gamma}_i$, if k_{ei} satisfies $k_{ei} - \frac{1}{2\tilde{\gamma}_i^2} - \frac{1}{2} > 0$, then the system (8) and (9) with input $V_{wo_i}^b$ and output e_{ei} has L_2-gain smaller than $\tilde{\gamma}_i$.

Item (i) shows that the visual motion observer leads the estimate $\bar{g}_{io_i}$ to the actual relative pose for a static object. Item (ii) implies that the observer works even for a moving target object, and the parameter $\tilde{\gamma}_i$ is an index on estimation accuracy.

IV. Passivity-based Cooperative Estimation

In this section, we present a cooperative estimation algorithm under the assumption of (i) each vision camera knows relative pose $g_{ij} = g_{wi}^{-1} g_{wj}$ with respect to the neighbors $j \in \mathcal{N}_i$ and (ii) $V_{wi}^b = 0 \ \forall i \in \mathcal{V}$. Under $V_{wi}^b = 0$, the relative rigid body motion (2) is simply given by $\dot{g}_{io_i} = g_{io_i} \hat{V}_{wo_i}^b$. Accordingly, the update procedure of the estimates in the visual motion observer is formulated as

$$\dot{\bar{g}}_{io_i} = \bar{g}_{io_i} \hat{u}_{ei}, \ u_{ei} = k_e e_{ei}, \tag{10}$$

where we restrict the gain k_{ei} as $k_{ei} = k_{ej} = k_e \ \forall i, j \in \mathcal{V}$.

A. Update Rule of Estimates

In this section, we present an update rule of the estimates $\bar{g}_{io_i}$ so as to estimate the mean g_i^* based on the visual motion observer [7] and passivity-based pose synchronization law [11]. For this purpose, we first assume that each vision camera i gains the estimate $\bar{g}_{jo_j}$ from $j \in \mathcal{N}_i$ as messages. Now, by multiplying known information g_{ij} from left, each vision camera i gets $\bar{g}_{io_j} := g_{ij}\bar{g}_{jo_j}$ for all $j \in \mathcal{N}_i$.

Let us define the update procedure of the estimate $\bar{g}_{io_i}$

$$\dot{\bar{g}}_{io_i} = \bar{g}_{io_i}\hat{u}_{ei},\ u_{ei} = k_e e_{ei} + \sum_{j\in\mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1}\bar{g}_{io_j}). \tag{11}$$

Since e_{ei} is reconstructed from the visual measurement f_i by (7) and $\bar{g}_{io_j}$ is obtained through communication as stated above, the update procedure (11) is implementable. The block diagram of the total cooperative estimation algorithm of vision camera i is illustrated in Fig. 4.

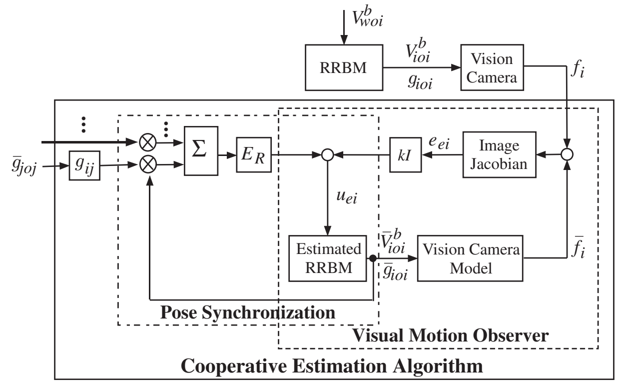

Fig. 4. Cooperative Estimation Algorithm

The present estimation algorithm is a kind of generalization of [7] and [10]. Indeed, as depicted in Fig. 4, without the second term, the update rule (11) is the same as that of the visual motion observer (10). This implies that the present algorithm structurally embodies the tracking property for moving targets. In addition, without the first term $k_e e_{ei}$, the update procedure (11), namely $u_{ei} = \sum_{j\in\mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1}\bar{g}_{io_j})$, is essentially equal to the passivity-based pose synchronization law [10] of a group of rigid bodies whose poses are represented by $\bar{g}_{wo_i} := g_{wi}\bar{g}_{io_i}$. Thus, under appropriate assumptions, each $\bar{g}_{wo_i}$ would converge to a state satisfying $\bar{g}_{wo_i} = \bar{g}_{wo_j}$ as time goes to infinity without the first term.

In the following, we mainly focus on the orientation part

$$\dot{e}^{\hat{\bar{\xi}}\bar{\theta}_{io_i}} = e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}\hat{\omega}_{uei}, \tag{12}$$

$$\omega_{uei} = k_e e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_{io_i}}) + \sum_{j\in\mathcal{N}_i} e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{io_j}}), \tag{13}$$

of (11) and closeness of the ultimate estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ to the average $e^{\hat{\xi}\theta_i^*}$.

B. Auxiliary Results

In this subsection, we give some auxiliary results necessary for proving the main result of this paper. Hereafter we use the following assumption.

Assumption 2: There exists a pair $(i,j) \in \mathcal{V}\times\mathcal{V}$ such that $e^{\hat{\xi}\theta_{wo_i}} \neq e^{\hat{\xi}\theta_{wo_j}}$ and $e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}} > 0$ for all $i \in \mathcal{V}$.

We first have the following lemma for the update procedure (12) and (13).

Lemma 1: Suppose that $V_{wo_i}^b = 0\ \forall i \in \mathcal{V}$ and the estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ are updated according to (12) and (13). Then, under Assumptions 1 and 2, if $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0\ \forall t \geq 0$, there exists a finite T such that each estimate $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ satisfies

$$\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \leq \sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}})\ \forall t \geq T. \tag{14}$$

Since the visual motion observer correctly estimates the pose $e^{\hat{\xi}\theta_{io_i}}$ without communication (Fact 1(i)), the right-hand side of (14) is equal to the sum of the energy function of the errors between the average $e^{\hat{\xi}\theta_i^*}$ and the ultimate estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ without communication, i.e. the estimation accuracy of the mean without communication. Thus, (14) means that the accuracy of the estimates as a group improves by using communication and cooperation. Though checking $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0$ for all time $t \geq 0$ is not easy, the subsequent Lemma 2 gives a solution to the problem.

Lemma 1 does not imply that each estimate $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ becomes close to $e^{\hat{\xi}\theta_i^*}$. In terms of the individual estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$, we get the following result.

Lemma 2: Suppose that $V_{wo_i}^b = 0\ \forall i \in \mathcal{V}$ and the estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ are updated according to (12) and (13). Then, under Assumption 1 and $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0\ \forall t \geq 0$, there exists a finite $T_1(c)$ satisfying

$$\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \leq \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}}) + c\ \forall t \geq T_1(c),\ i \in \mathcal{V}, \tag{15}$$

for any scalar $c > 0$, where $h := \arg\max_j \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_j}})$. This lemma implies that individual estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ get closer to the mean $e^{\hat{\xi}\theta_i^*}$ at least than the object with the farthest orientation from the mean. In addition, the proof of this lemma also means that the set $\mathcal{S} = \{(e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})_{i\in\mathcal{V}}|\ e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0\ \forall i \in \mathcal{V}\}$ is positively invariant for (12) and (13) under Assumption 2 if the object is static. Namely, if $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0$ is satisfied at the initial time and Assumption 2 is true, then $e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}}e^{\hat{\xi}\theta_i^*} > 0$ holds for all subsequent time and we do not need to check it in transient states. Hereafter, we use the notation $\delta := \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}})$, $\delta_c = \delta + c$ for notational simplicity. Note that these parameters are fixed values.

Let us now define the following subsets of $\mathcal{S}$ for some $\varepsilon \in [0,1]$, where $\rho := \sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}})$ and $\beta := 1 - \sqrt{2\delta_c}$.

$$\mathcal{S}_1 := \Big\{(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}\Big|\ \sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) > \rho\Big\},$$

$$\mathcal{S}_2(k_e) := \Big\{(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}\setminus\mathcal{S}_1\Big| \sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{N}_i}\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \geq \frac{k_e\rho}{\beta}\Big\}$$

$$\mathcal{S}_3(k_e,\varepsilon) := \Big\{(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}\setminus(\mathcal{S}_1\cup\mathcal{S}_2(k_e))\Big|$$

$$\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) > \rho\varepsilon\Big\}.$$

Then, Lemma 1 says that the function $V = \sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})$ is strictly decreasing in the set $\mathcal{S}_1$ along with the trajectories of (12) and (13). In the other regions, we obtain the following lemma.

Lemma 3: Suppose that all the assumptions in Lemma 1 hold and $\beta > 0$. Then, the function V is strictly decreasing in the region $\mathcal{S}_2(k_e)$. In addition, in the region $\mathcal{S}_3(k_e,\varepsilon)$, there exists an $i\in\mathcal{V}$ such that

$$\phi(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}) \le k_e\frac{\mathrm{diam}(G)^2}{\beta}\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_i}}) \quad (16)$$

for all $j\in\mathcal{V}$, where $\mathrm{diam}(G)$ is the diameter of the graph G [9].

Note that the assumption $\beta > 0$ means that the absolute value of the rotation angle of $e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_i}}$ is smaller than about 41[deg] for all $i\in\mathcal{V}$.

C. Main Result

Now, we are ready to state the main result of this paper which gives a quantitative evaluation on the closeness of the ultimate estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}$ to the mean $e^{\hat{\xi}\theta_i^*}$.

Theorem 1: Suppose that all the assumptions in Lemma 1 hold. Then, for any $\epsilon\in(0,1)$, there exists a finite $T(\epsilon) > 0$ such that

$$\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \le \alpha\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_i}})\ \forall t\ge T(\epsilon), \quad (17)$$

where α is a function of β and $k_e > 0$ given by

$$\alpha(\beta,k_e) = 1-(1-\epsilon)\left(\sqrt{\beta}-\sqrt{k_e n}\mathrm{diam}(G)\right)^2 \quad (18)$$

for positive $\beta > 0$ and k_e satisfying

$$k_e \le \frac{\beta}{n\mathrm{diam}(G)^2}. \quad (19)$$

and otherwise $\alpha(\beta,k_e)=1$.

Inequality (17) says that if the parameter α is small enough, the mean estimation becomes accurate. Under (19), α gets small as the term $\sqrt{k_e n}\mathrm{diam}(G)$ approaches to 0. Now, if we use a sufficiently small gain k_e in (11), then the term is approximated by 0. In addition, if we take ϵ sufficiently close to 0, then (18) is approximately given by $1-\beta=\sqrt{2\delta_c}$. From the definition of $\delta_c := \phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta_{wo_h}})+c$, if the target object's orientation $e^{\hat{\xi}\theta_{wo_h}}$ is sufficiently close to the mean $e^{-\hat{\xi}\theta^*}$, i.e. if $e^{\hat{\xi}\theta_{wo_i}}$ and $e^{\hat{\xi}\theta_{wo_j}}$ are close enough among all $i,j\in\mathcal{V}$, then it becomes close to 0 and the mean is accurately estimated by the present update procedure (12) and (13). Otherwise, the accuracy might degrade, though it is more accurate at least than the case in the absence of cooperation. The parameter ϵ is associated with the speed of convergence $T(\epsilon)$ and for a small ϵ the time to take for reaching the region satisfying (17) becomes long.

D. Relation to Multi-agent Optimization

In this subsection, we give a relation between the above result and the multi-agent optimization technique in [16]. The multi-agent optimization problem is formulated as

$$\min_{x\in\mathbb{R}^n} F(x) := \sum_{i=1}^{N}F_i(x)\ (F_i:\text{convex}), \quad (20)$$

where agent i does not know any $F_j,\ j\neq i$. For the problem, [16] presents the update rule of the estimate of solution $x_i[t]$, $x_i[t+1] = -k_e\mathrm{grad}_{x_i[t]}F_i+\sum_{j\in\mathcal{V}}a_{ij}x_j[t]$, where $\mathrm{grad}_{x_i[t]}F_i$ is a gradient of F_i at $x_i[t]$, $a_{ij}=a_{ji},\ a_{ij}=0$ if $(i,j)\notin\mathcal{E}$ and $\sum_{j\in\mathcal{V}}a_{ij}=1\ \forall i\in\mathcal{V}$. This rule consists of the consensus [8] and gradient descent of the individual objective function F_i. [16] also derives an upper bound $\varepsilon_{t,i}$ of the error between $F(x^*)$ and $F(x_i[t])$, where x^* is an actual optimal solution to (20). In addition, they show that an upper bound for ultimate estimates, $\varepsilon_i := \lim_{t\to\infty}\epsilon_{t,i}$, can be arbitrarily small as k_e becomes small.

Now, we consider the multi-agent optimization on $SO(3)$

$$\min_{e^{\hat{\xi}\theta}\in SO(3)}\sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta}e^{\hat{\xi}\theta_{wo_i}}). \quad (21)$$

The solution to (21) is given by the Euclidean mean $e^{\hat{\xi}\theta^*}$. If the update procedure (12) and (13) is viewed as that of $e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}$ in the world frame, we get

$$\begin{aligned}\dot{e}^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} &= k_e e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\xi}\theta_{wo_i}}) \\ &\quad + e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}\sum_{j\in\mathcal{N}_i}e_R(e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}}e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}). \quad (22)\end{aligned}$$

The first term of (22) gives the gradient decent of the individual objective function $\phi(e^{-\hat{\xi}\theta}e^{\hat{\xi}\theta_{wo_i}})$ on $SO(3)$ [18]. Since the consensus algorithm [8] is not implementable on $SO(3)$, we instead use the attitude synchronization law [11] achieving convergence of the orientations to a common value though the convergence value is not equal to the average of the initial orientations.

The conclusion in Subsection IV-C basically corresponds to [16], i.e. the estimates get close to the mean as k_e becomes small. However, unlike problems on a vector space as in [16], the present scheme yields an offset of the estimate from the average regardless of the gain k_e. This is reasonable from the fact that the attitude synchronization law does not always drive the orientations to the average of the initial values.

V. Experiments

Finally, we demonstrate the effectiveness of the present algorithm through experiments with three vision cameras. For detailed information on the experimental environment, the readers are recommended to refer to [14]. Let the camera poses relative to the world frame be set as $p_{w1} = [-0.3\ 0.18\ 0.00]^T$, $p_{w2}=0$, $p_{w3} = [0.22\ 0.18\ 0.00]^T$, $\xi\theta_{w1} = [0.00\ 0.38\ 0.00]^T$, $\xi\theta_{w2} = 0$, $\xi\theta_{w3} = [0.00\ -0.44\ 0.00]^T$. We also assume the communication graph $\mathcal{E} = \{(1,2),(2,1),(1,3),(3,1)\}$.

We set a single real target object. Even for a single object, the visual motion observer gives different estimates

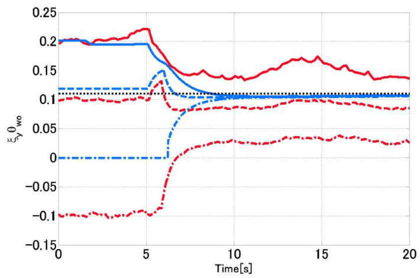

Fig. 5. Time Responses of Estimates

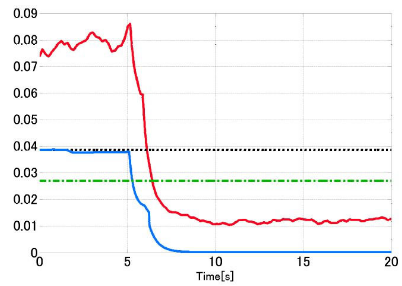

Fig. 6. Time Responses of Objective Function

among cameras due to the distortion of lenses or parametric uncertainties of cameras. A way to improve estimation accuracy under such uncertainties is averaging the poses, namely the problem is reduced to that under consideration in this paper by letting the estimates without communication be g_{io_i}. In this experiment, the estimates without communication converge to the poses $p_{wo_1} = [0.05\ 0.16\ 0.70]^T$, $p_{wo_2} = [0.07\ 0.13\ 0.74]^T$, $p_{wo_3} = [0.01\ 0.14\ 0.76]^T$, $\xi\theta_{wo_1} = [-0.11\ 0.20\ -0.01]^T$, $\xi\theta_{wo_2} = [0.07\ 0.12\ -0.03]^T$, $\xi\theta_{wo_3} = [0.00\ 0.00\ 0.00]^T$. The Euclidean mean $e^{\hat{\xi}\theta^*}$ of the orientations $\{e^{\hat{\xi}\theta_{wo_i}}\}_{i\in\{1,2,3\}}$ is given by $\xi\theta^* = [-0.01\ 0.11\ -0.02]^T$. Then, the parameter β in Theorem 1 is given by $\beta = 0.8 > 0$ for $c = 10^{-4}$.

We run the present algorithm for two different gains $k_e = 0.01$ and $k_e = 0.8$, where the former satisfies (19) and the latter does not. For $k_e = 0.01$, α in (18) is given by 0.70 with $\epsilon = 10^{-4}$ and for $k_e = 0.8$ it is 1. Though we cannot show all the responses due to page constraints, Fig. 5 shows the time responses of estimates and Fig. 6 those of the objective function, where blue curves are the responses with $k_e = 0.01$ and red ones with $k_e = 0.8$. In Fig. 5, the solid, dashed and dotted curves illustrate the estimates of camera 1, 2 and 3 respectively. We see from the figure that a small gain k_e gives more accurate estimate of the mean $(\xi\theta^*)_2 = 0.11$ than a large gain, which validates the statement just after Theorem 1. In Fig. 6, the black line shows the value of $\sum_{i=1}^{3}\phi(e^{-\hat{\xi}\theta^*}e^{\hat{\xi}\theta}_{wo_i})$ and the green one is the right hand side of (18) with $k_e = 0.01$. We see from the figure that the validity of Theorem 1 is demonstrated, i.e. the inequality (18) is really satisfied for both gains. However, the case of $k_e = 0.8$ also gets lower than the green line, which indicates conservatism of Theorem 1 and its reduction should be tackled in the future.

VI. Conclusions

This paper has presented a novel cooperative estimation algorithm for visual sensor networks. We have considered the situation where multiple smart vision cameras with computation and communication capability see different target objects. We first have presented an estimation algorithm to meet two requirements, averaging and tracking. Then, we have provided an upper bound of the ultimate error between the actual average and the estimates given by the present algorithm. Finally the effectiveness of the present estimation algorithm has been demonstrated through experiments.

References

[1] H. Aghajan and A. Cavallaro (Eds), "Multi-Camera Networks: Principles and Applications," Academic Press, 2009.

[2] M. Zhu and S. Martinez, "Distributed Coverage Games for Mobile Visual Sensors (I), Reaching the set of Nash equilibria," Proc. of the 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, pp. 169–174, 2009.

[3] G. Chesi and K. Hashimoto (Eds), "Visual Servoing via Advanced Numerical Methods," Lecture Notes in Control and Information Sciences, Vol. 401, Springer-Verlag, 2010.

[4] M. Johnson, G. Hu, K. Dupree, W. E. Dixon, "Inverse Optimal Homography-based Visual Servo Control via An Uncalibrated Camera," Proc. of the 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, pp. 2408-2414, 2009.

[5] M. Sznaier, M. Ayazoglu, O. I. Camps, "Using Dynamics to Recover Euclidian 3-dimensional Structure from 2-dimensional Perspective Projections," Proc. of the 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, pp. 2414-2419, 2009.

[6] A. Censi, S. Han, S. B. Fuller, and R. M. Murray, "A Bio-Plausible Design for Visual Attitude Stabilization," Proc. of 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, pp. 3513-3520, 2009.

[7] M. Fujita, H. Kawai and M. W. Spong, "Passivity-based Dynamic Visual Feedback Control for Three Dimensional Target Tracking:Stability and L2-gain Performance Analysis," IEEE Trans. on Control Systems Technology, Vol.15, No.1, pp.40-52, 2007.

[8] R. Olfati-Saber, J. A. Fax and R. M. Murray, "Consensus and Cooperation in Networked Multi-Agent Systems," Proc. of the IEEE, Vol. 95, No. 1, pp. 215–233, 2007.

[9] F. Bullo, J. Cortes and S. Martinez, "Distributed Control of Robotic Networks," Princeton Series in Applied Mathematics, 2009.

[10] Y. Igarashi, T. Hatanaka, M. Fujita and M. W. Spong, "Passivity-based Output Synchronization and Flocking Algorithm in SE(3)," Proc. of the 47th IEEE Conference on Decision and Control, pp. 1024-1029, 2008

[11] Y. Igarashi, T. Hatanaka, M. Fujita and M. W. Spong, "Passivity-based Attitude Synchronization in $SE(3)$," IEEE Trans. on Control Systems Technology, Vol. 17, No. 5, pp. 1119–1134, 2009.

[12] R. Olfati-Saber, "Distributed Kalman Filter for Sensor Networks," Proc. of the 46th IEEE Conference on Decision and Control, pp.5492-5498, 2007.

[13] R. Tron, R. Vidal and A. Terzis, "Distributed Pose Averaging in Camera Sensor Networks via Consensus on SE(3)," Proc. of the International Conference on Distributed Smart Cameras, 2008.

[14] T. Hatanaka, M. Fujita and F. Bullo "Vision-based Cooperative Estimation via Multi-agent Optimization," Proc. of the 49th IEEE Conference on Decision and Control, to appear, 2010.

[15] M. Moakher, "Means and Averaging in the Group of Rotations," SIAM Journal on Matrix Analysis and Applications, Vol. 24, No. 1, pp. 1–16, 2002.

[16] A. Nedic and A. Ozdaglar, "Distributed Subgradient Methods for Multi-agent Optimization," IEEE Trans. on Automatic Control, Vol. 54, No. 1, pp. 48-61, 2009.

[17] Y. Ma, S. Soatto, J. Kosecka and S. S. Sastry, "An Invitation to 3-D Vision: From Images to Geometric Models," Springer, 2004.

[18] P. A. Absil, R. Mahony and R. Sepulchre, "Optimization Algorithms on Matrix Manifolds," Princeton Press, 2008.