2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Cooperative Estimation of 3D Target Object Motion via Networked Visual Motion Observers

Takeshi Hatanaka, Kenji Hirata and Masayuki Fujita

Abstract— **This paper addresses cooperative estimation of 3D target motion for visual sensor networks. In one of our previous works, we already presented a cooperative estimation algorithm called networked visual motion observers. In this paper, we first clarify averaging accuracy attained by the networked estimation mechanism. Then, we analyze convergence speed of the estimates and clarify a relation of the speed to the graph structures. Moreover, we also reveal a tracking performance of the estimates to target objects motion and derive a connection between the tracking performance and a visual feedback gain in the algorithm. Finally the effectiveness of the present estimation algorithm is demonstrated through experiments.**

I. Introduction

In this paper, we address estimation of 3D target object motion for visual sensor networks, which is a network consisting of spatially distributed smart cameras with communication and computation capability [1], [2]. The visual sensor networks include applications such as environmental monitoring, surveillance, target tracking and entertainment.

A lot of research works have been devoted to combining systems and control theory with vision [3]–[6] and we here focus on estimation of 3D rigid body motion as in [5], [6]. In visual sensor networks, it is expected that not only an estimate is produced but also the vision cameras cooperate with each other, which brings us new theoretical challenges. The advantages of the cooperation is (i) to improve estimation accuracy by integrating rich information, (ii) to gain tolerance against obstruction, misdetection and sensor failures and (iii) to eliminate blind areas by fusing images from a variety of viewpoints. In addition, to build scalable networks, it is required to achieve cooperation in a distributed fashion. To tackle such problems, cooperative control schemes as in [7]–[9] provide useful methodologies.

Cooperative estimation for sensor networks has been studied (e.g. [9], [10]), whose main objective is averaging the sensed data or local estimates in order to improve the estimation accuracy. However, most of the algorithms are not applicable to our problem since the object's pose takes values in a non-Euclidean space. Meanwhile, [11] presents a distributed estimation algorithm for visual sensor networks. However, they assume that the target's orientation is obtained *a priori* and do not mention estimation from vision data. In order to present an algorithm achieving estimation and averaging simultaneously, the authors [12] presented a cooperative estimation algorithm consisting of not only visual feedbacks but also mutual feedbacks from neighboring vision cameras based on the passivity-based visual motion observer [5], [6] and pose synchronization law [8]. In this paper, the resulting cooperative estimation mechanism is called a *networked visual motion observers*.

In this paper, we extend the result of [12] in several aspects. Especially, we give analysis on the networked visual motion observer in terms of averaging accuracy, convergence speed and tracking performance. After introducing the visual motion observer presented in [5], we present the networked visual motion observer originally given in [12]. We next formally define the notion of averaging accuracy and reveal the accuracy attained by the present mechanism, whose partial solution is already presented in [12] and we provide its generalized version. We then prove that the time to achieve the averaging accuracy is upper bounded by a function of the communication graph and feedback gains. The result clarifies that the convergence time becomes short if: (i) the maximal eigenvalue of the graph Laplacian [7] is close to the second smallest one called algebraic connectivity and (ii) the visual feedback gain is large. Then, we also tackle the tracking of the estimates to the average for moving target objects. The result therein gives us an insight that the tracking performance improves as the visual feedback gain increases. We finally show the effectiveness of the algorithm through experiments on a testbed of visual sensor networks.

In this paper, we use the following notations. The readers are recommended to refer to [3] for more details on the terminologies. We use the notation $e^{\hat{\xi}_{ab}\theta_{ab}} \in \mathcal{R}^{3\times3}$ 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.

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
Kenji Hirata is with the Department of Mechanical Engineering, Nagaoka University of Technology, Nagaoka 940 2188, JAPAN hirata@nagaokaut.ac.jp

978-1-61284-799-3/11/$26.00 ©2011 IEEE

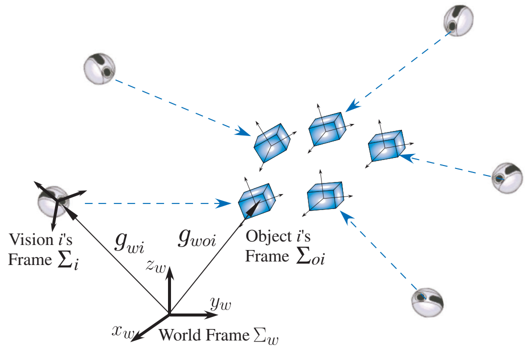

Fig. 1. Visual Sensor Networks

II. Preliminaries

Throughout this paper, we consider the situation where n vision cameras $i \in \mathcal{V} := \{1, \cdots, n\}$ see a group of target objects $o_i,\ i \in \mathcal{V}$ (Fig. 1). Each vision camera i captures the object o_i on its image plane.

A. Rigid Body Motion

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 o_i, respectively. Then, the pose of vision camera i and object o_i relative to Σ_w are denoted by $g_{wi} = (p_{wi}, e^{\hat{\xi}\theta_{wi}}) \in SE(3)$ and $g_{wo_i} = (p_{wo_i}, e^{\hat{\xi}\theta_{wo_i}}) \in SE(3)$. The pose of Σ_{o_i} relative to Σ_i is represented by $g_{io_i} = (p_{io_i}, e^{\hat{\xi}\theta_{io_i}}) \in SE(3)$, which is given by $g_{io_i} = g_{wi}^{-1} g_{wo_i}$.

We next define the body velocity of the object Σ_{o_i} relative to the world frame Σ_w as $V_{wo_i}^b = (v_{wo_i}, \omega_{wo_i}) = (g_{wo_i}^{-1}\dot{g}_{wo_i})^\vee \in \mathcal{R}^6$, where v_{wo_i} and ω_{wo_i} represent the linear and the angular velocity of the origin of Σ_{o_i} relative to Σ_w, respectively [3]. Similarly, vision camera i's body velocity relative to Σ_w will be denoted as $V_{wi}^b = (v_{wi}, \omega_{wi}) = (g_{wi}^{-1}\dot{g}_{wi})^\vee$. Then, the motion of the relative pose g_{io_i} is represented by

$$\dot{g}_{io_i} = -\hat{V}_{wi}^b g_{io_i} + g_{io_i}\hat{V}_{wo_i}^b. \tag{1}$$

Equation (1) is a standard formula for the relation among the body velocities of three coordinate frames [3].

B. Visual Measurement

In this subsection, we define the visual measurement of vision camera i which is available for estimation of target object motion. Throughout this paper, we use the pinhole camera model with a perspective projection [3].

We assume that target object o_i has m feature points and vision camera i can extract them from the 2D visual data by using some techniques. The position vectors of the target object o_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 $[p_{o_il}^T\ 1]^T$ and $[p_{il}^T\ 1]^T$.

Let the m feature points of the object o_i on the image plane coordinate, denoted by $f_i := [f_{i1}^T \cdots f_{im}^T]^T \in \mathcal{R}^{2m}$, be the measurement of the camera i. It is well known [3] 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

$$f_{il} = \lambda_i \left[x_{il}/z_{il} \quad y_{il}/z_{il}\right]^T,\ p_{il} = [x_{il}\ y_{il}\ z_{il}]^T \tag{2}$$

where λ_i is a focal length of camera i.

C. Visual Motion Observer

In this subsection, we consider the problem that a vision camera i estimates the target object motion g_{io_i} from the visual measurement f_i. For this purpose, we introduce the visual motion observer presented in [5].

We first prepare a model of the actual relative rigid body motion (1) as

$$\dot{\bar{g}}_{io_i} = -\bar{g}_{io_i}\hat{V}_{wi}^b + \bar{g}_{io_i}\hat{u}_{ei}. \tag{3}$$

where $\bar{g}_{io_i} = (\bar{p}_{io_i}, e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \in SE(3)$ is the estimate of the actual relative pose g_{io_i}. The input $u_{ei} = (v_{uei},\ \omega_{uei})$ is to be determined to drive the estimated values $\bar{g}_{io_i}$ to the actual value g_{io_i}. Once the estimate $\bar{g}_{io_i}$ is determined, the estimated measurement $\bar{f}_i$ is also computed by (2).

We next 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

$$\begin{aligned} E_R(g) &:= [p^T\ e_R^T(e^{\hat{\xi}\theta})]^T,\ g = (p, e^{\hat{\xi}\theta}),\\ e_R(e^{\hat{\xi}\theta}) &:= \mathrm{sk}(e^{\hat{\xi}\theta})^\vee,\ \mathrm{sk}(e^{\hat{\xi}\theta}) := \frac{1}{2}(e^{\hat{\xi}\theta} - e^{-\hat{\xi}\theta}), \end{aligned}$$

the vector representation of the estimation error is defined by $e_{ei} := E_R(g_{ei})$.

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 $f_{ei} = J_i(\bar{g}_{io_i})e_{ei}$ [5], 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 e_{ei} is reconstructed from visual measurement f_i and $\bar{g}_{io_i}$ as

$$e_{ei} = J_i^\dagger(\bar{g}_{io_i})f_{ei}, \tag{4}$$

where $\dagger$ denotes the pseudo-inverse.

Based on the fact that the estimation error system

$$\dot{g}_{ei} = -\hat{u}_{ei}g_{ei} + g_{ei}\hat{V}_{wo_i}^b \tag{5}$$

is passive from u_{ei} to $-e_{ei}$ if $V_{wo_i}^b = 0$, the input

$$u_{ei} = -k_e(-e_{ei}) = k_e e_{ei},\ k_e > 0 \tag{6}$$

is presented in [5]. The total estimation mechanism (3), (4) and (6) is called *visual motion observer* [6].

It is also shown in [5] from the passivity-based control theory that in case of $V_{wo_i}^b = 0$ the equilibrium point $e_{ei} = 0$ for the closed-loop system (5) and (6) is asymptotically stable, which implies that the visual motion observer leads the estimate $\bar{g}_{io_i}$ to the actual relative pose for a static object. Moreover, the tracking performance of the estimate to the target object motion g_{io_i} is analyzed in the framework of the L_2-gain analysis.

III. Networked Visual Motion Observers

In the subsequent sections, we consider the group of vision cameras $\mathcal{V}$ together with the group of the target objects $\{o_i\}_{i\in\mathcal{V}}$ as in Fig. 1. We assume that the vision cameras can communicate with each other and some information of the neighboring cameras is available for updating its estimate $\bar{g}_{io_i}$. The communication network is modeled by a graph $G = (\mathcal{V}, \mathcal{E})$, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ as in [7]. In addition, we define the neighbor set $\mathcal{N}_i := \{j \in \mathcal{V} |\ (j, i) \in \mathcal{E}\}$.

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

A. Averages on $SO(3)$ and $SE(3)$

The objective of this paper is to achieve averaging for static objects, which means the estimates $\bar{g}_{wo_i} := g_{wi}\bar{g}_{io_i}$ become close to an average of $(g_{wo_i})_{i\in\mathcal{V}}$, while preserving the tracking nature of the visual motion observer for moving target objects. The problem is motivated by estimation of a single object motion under uncertain measurements and estimation of multiple objects motion behaving as a group. However, in terms of the latter scenario, more thorough investigations, e.g. on segmentation of objects in the image and allocation of targets to each camera, are necessary.

Let us introduce an average g^* of $\{g_{wo_j}\}_{j\in\mathcal{V}}$ as

$$g^* = (p^*, e^{\hat{\xi}\theta^*}) := \arg \min_{g\in SE(3)} \sum_{i\in\mathcal{V}} \psi(g^{-1}g_{wo_i}), \qquad (7)$$

$$\psi(g) := \frac{1}{2}\|I_4 - g\|_F^2 = \frac{1}{2}\|p\|^2 + \phi(e^{\hat{\xi}\theta}),\ g = (p, e^{\hat{\xi}\theta})$$

$$\phi(e^{\hat{\xi}\theta}) := \frac{1}{2}\|I_3 - e^{\hat{\xi}\theta}\|_F^2 = \mathrm{tr}(I_3 - e^{\hat{\xi}\theta}),$$

where $\|M\|_F$ is the Frobenius norm of matrix M. We also use the notation $g_i^* = (p_i^*, e^{\hat{\xi}\theta_i^*}) := g_{wi}^{-1}g^*$. The position average p^* is equal to the arithmetic mean $p^* = \frac{1}{n}\sum_{j\in\mathcal{V}} p_{wo_j}$ of $\{p_{wo_j}\}_{j\in\mathcal{V}}$ and the orientation average $e^{\hat{\xi}\theta^*}$ is a so-called Euclidean mean [13] of $\{e^{\hat{\xi}\theta_{wo_j}}\}_{j\in\mathcal{V}}$. It is known [13] that the Euclidean mean $e^{\hat{\xi}\theta^*}$ is given by

$$e^{\hat{\xi}\theta^*}(t) = \mathrm{Proj}(S(t)),\ S(t) := \frac{1}{n}\sum_{j\in\mathcal{V}} e^{\hat{\xi}\theta_{wo_j}}(t). \qquad (8)$$

Here, $\mathrm{Proj}(M)$ gives the orthogonal projection of M onto $SO(3)$, which is given by $U_M V_M^T$ for the matrix M with singular value decomposition $M = U_M \Sigma_M V_M^T$.

B. Networked Visual Motion Observers

In this subsection, we introduce a cooperative estimation algorithm under the assumption that each vision camera is static $V_{wi}^b = 0$ and knows relative pose $g_{ij} = g_{wi}^{-1}g_{wj}$ with respect to the neighbors $j \in \mathcal{N}_i$.

Each vision camera i first gains the estimates $\bar{g}_{jo_j}$ from $j \in \mathcal{N}_i$ as messages. Then, by multiplying known 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 now define the update procedure of the estimate $\bar{g}_{io_i}$ as (3) with $V_{wi}^b = 0$ and

$$u_{ei} = k_e e_{ei} + k_s \sum_{j\in\mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1}\bar{g}_{io_j}), \qquad (9)$$

where $k_e > 0, k_s > 0$. Note that e_{ei} is reconstructed from (4) and $\bar{g}_{io_j}$ is obtained through communication as stated above. Thus, the procedure (9) is implementable from the visual measurement (2). The block diagram of the total system associated with vision camera i is illustrated in Fig. 2.

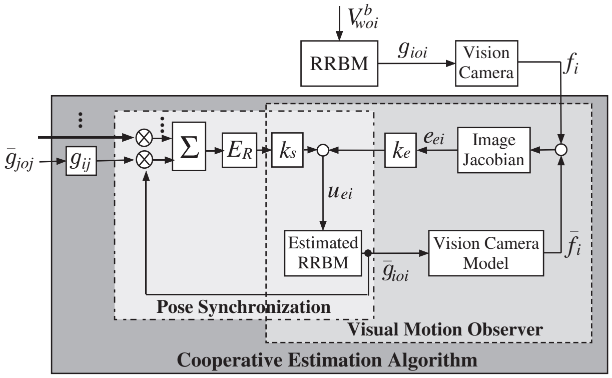

Fig. 2. Cooperative Estimation Algorithm

The present algorithm (9) consists of the visual feedback $k_e e_{ei}$ and the mutual feedback $k_s \sum_{j\in\mathcal{N}_i} E_R(\bar{g}_{io_i}^{-1}\bar{g}_{io_j})$. As depicted in Fig. 2, without the second term, the update rule (9) is the same as that of the visual motion observer (6). The form of the mutual feedback is inspired by the pose synchronization law [8] of a group of rigid bodies and indeed, without the first term $k_e e_{ei}$, the update rule (9) is essentially equal to the law in [8]. In other words, the visual motion observers are networked by the mutual feedback term in the total estimation mechanism (3), (4) and (9). We thus call the mechanism *networked visual motion observers*.

IV. Averaging Accuracy and Convergence Speed

In this section, we investigate estimation accuracy of the average g_i^* for the the networked visual motion observers and convergence speed under the following assumption.

Assumption 2: (i) The target objects are static, i.e. $V_{wo_i}^b = 0\ \forall i \in \mathcal{V}$. (ii) There exists a pair $(i, j) \in \mathcal{V}\times\mathcal{V}$ such that $e^{\hat{\xi}\theta_{wo_i}} \neq e^{\hat{\xi}\theta_{wo_j}}$. (iii) $e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}} > 0$ for all $i \in \mathcal{V}$.
In terms of the item (iii), we have

$$\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}}) \leq \phi_m := \max_{i,j\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_{wo_i}}e^{\hat{\xi}\theta_{wo_j}})\ \forall i \in \mathcal{V}\ (10)$$

as long as $e^{-\hat{\xi}\theta_{wo_i}}e^{\hat{\xi}\theta_{wo_j}} > 0\ \forall i, j \in \mathcal{V}$, though we omit its proof. (10) implies that if ϕ_m is smaller than 2, then $\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}}) \leq 2\ \forall i \in \mathcal{V}$ and hence (iii) is satisfied. Thus, (iii) can be checked if set-valued prior information on the target orientations, i.e. an upper value of ϕ_m is available.

A. Analysis on Averaging Accuracy

In this subsection, we introduce a notion of approximate averaging. Due to the page constraints, we focus only on the evolution of the orientation estimates. If we extract only the orientation part from (3) and (9), we obtain

$$\dot{e}^{\hat{\bar{\xi}}\bar{\theta}_{io_i}} = e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}\Big(k_e \mathrm{sk}(e^{\hat{\xi}\theta_{ei}}) + k_s \sum_{j\in\mathcal{N}_i} \mathrm{sk}(e^{\hat{\bar{\xi}}\bar{\theta}_{o_{ij}}})\Big) \qquad (11)$$

with $e^{\hat{\bar{\xi}}\bar{\theta}_{o_{ij}}} := e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{io_j}} = e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_j}}$, which is independent of the evolution of position estimates.

Let us now define *ε-level averaging accuracy* as below.

Definition 1: Given target object poses $(g_{io_i})_{i\in\mathcal{V}}$, the estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ achieve ε-level averaging accuracy if there exists a finite T such that

$$(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t))_{i\in\mathcal{V}} \in \Omega(\varepsilon)\ \forall t \geq T,\ \rho := \sum_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*} e^{\hat{\xi}\theta_{io_i}}).$$

$$\Omega(\varepsilon) := \left\{ (e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \,\middle|\, \sum_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*} e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \leq \varepsilon\rho \right\}, \qquad (12)$$

In the absence of communication, what each vision camera can do is to provide as an accurate estimate of g_{io_i} as possible. Namely, the parameter ρ indicates the best performance of average estimation in the absence of communication. More specifically, the visual motion observer correctly estimates g_{io_i} if $V^b_{wo_i} = 0$ and hence ρ indicates the ultimate estimation accuracy of the average in the absence of the mutual feedbacks and the parameter ε is an indicator of improvement of average estimation accuracy by inserting the mutual feedbacks.

In terms of the averaging accuracy by the networked visual motion observers, we have the following result, where $k = k_e/k_s$, $\delta := \max_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\xi}\theta_{wo_i}})$, $\delta_c := \delta + c$ with a positive scalar c, $\beta := 1 - \sqrt{2\delta_c}$ and the set $\mathcal{S}$ is defined by $\mathcal{S} := \{(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} |\ e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}} e^{\hat{\xi}\theta_i^*} > 0\ \forall i \in \mathcal{V}\}$.

Theorem 1: Suppose that the estimates $(\bar{g}_{io_i})_{i\in\mathcal{V}}$ are updated by the networked visual motion observer (3), (4) and (9). Under Assumptions 1 and 2, if the initial estimates satisfy $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(0))_{i\in\mathcal{V}} \in \mathcal{S}$, for any $\epsilon \in (0,1)$ and $c > 0$, the orientation estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ achieve ε_{ave}-level averaging accuracy with

$$\varepsilon_{ave} = \begin{cases} 1 - (1-\epsilon)\left(\sqrt{\beta} - \sqrt{k}d_m\right)^2, & \\ \qquad\qquad\qquad\qquad \text{if } k \leq \beta/d_m^2,\ \beta > 0 & \\ 1, \text{ otherwise} & \end{cases}, \qquad (13)$$

where d_m is defined by $d_m = \min_{j\in\mathcal{V}} \sqrt{\sum_{i\in\mathcal{V}} d_{ij}^2}$ and d_{ij} is the size of the shortest path along with the graph G whose edges are replaced by undirected ones.

Equation (13) says that if we choose a sufficiently small k, i.e. $k_s \gg k_e$ in (9), ε_{ave} becomes small and the average estimation accuracy improves. Note that the parameter δ is upper bounded by ϕ_m and a lower bound of β is derived if set valued prior information on target orientations are available.

The proof for a special case with $k_s = 1$ is already given in [12] and the above theorem is its generalized version. However, the procedure of the proof is almost the same as [12] and we show only the sketch of the proof in the next subsection omitting the details.

B. Sketch of Proof of Theorem 1

In this subsection, we briefly review the proof of Theorem 1. We first give the following lemma.

Lemma 1: [12] Suppose that all the assumptions of Theorem 1 hold. Then, for all $c > 0$, there exists finite $\tau(c)$ such that $\phi(e^{-\hat{\xi}\theta_i^*} e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) \leq \delta_c$ for all $t > \tau(c)$. In addition, $\tau(c)$ is upper bounded by

$$\bar{\tau}(c) := \frac{1}{c}(\max\{0, \max_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*} e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(0)) - \delta\}) + 1.$$

From this lemma, it is easily proved that the set $\mathcal{S}$ is positively invariant for (3) with (9) under Assumption 2 [12].

We next define the following sets.

$$\mathcal{S}_0 := \mathcal{S} \cap \Omega(1),\ \mathcal{S}_1 := \mathcal{S} \setminus \mathcal{S}_0$$
$$\mathcal{S}_2(k) := \left\{ (e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}_0 \middle| \beta \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} \phi(e^{\hat{\bar{\xi}}\bar{\theta}_{o_{ij}}}) \geq k\rho \right\}$$
$$\mathcal{S}_3(k,\varepsilon) := \mathcal{S}_0 \setminus (\mathcal{S}_2(k) \cup \Omega(\varepsilon))$$

for some $\varepsilon \in [0,1)$. It should be now noted that

$$\mathcal{S}_0 \setminus (\mathcal{S}_2(k) \cup \mathcal{S}_3(k,\varepsilon)) \subseteq \Omega(\varepsilon). \qquad (14)$$

In the proof of Theorem 1, we employ the energy function

$$V := \sum_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_i^*} e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}) = \sum_{i\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}).$$

Then, if $k \leq \beta/d_m^2$ and $\beta > 0$, we have

$$\dot{V} \leq -a_j,\ \text{if } (e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}} \in \mathcal{S}_j, j = 1,2,3 \qquad (15)$$

at least after the time $\tau(c)$ under Assumptions 1 and 2, where

$$a_1 := \beta \sum_{i\in\mathcal{V}} \left(k_e\phi(e^{\hat{\xi}\theta_{ei}}) + k_s \sum_{j\in\mathcal{N}_i} \phi(e^{\hat{\bar{\xi}}\bar{\theta}_{o_{ij}}}) \right),$$
$$a_2 := \sum_{i\in\mathcal{V}} k_e \left(\phi(e^{-\hat{\xi}\theta^*} e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}) + \beta\phi(e^{\hat{\xi}\theta_{ei}}) \right),$$
$$a_3 := \beta \sum_{i\in\mathcal{V}} \left(k_e\epsilon\phi(e^{\hat{\xi}\theta_{ei}}) + k_s \sum_{j\in\mathcal{N}_i} \phi(e^{\hat{\bar{\xi}}\bar{\theta}_{o_{ij}}}) \right).$$

Under Assumptions 1 and 2, a_1, a_2 and a_3 are all strictly positive and hence the estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ settle into the set $\Omega(\varepsilon)$ in a finite time from (14). This completes the proof.

Let us define the parameters $\bar{a}_i = \min_{(e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}})_{i\in\mathcal{V}}} a_i,\ i = 1,2,3$. Then, (15) holds even if a_i is replaced by $\bar{a}_i$, and $\bar{a}_i$ can be measures of the convergence speed.

C. Convergence Speed Analysis

This paper employs the time to achieve ε-level averaging accuracy as an index to measure the convergence speed as

$$T_\varepsilon = \inf\{T \geq 0 |\ (e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t))_{i\in\mathcal{V}} \in \Omega(\varepsilon)\ \forall t \geq T\}.$$

The objective here is to derive an upper-bound of T_ε when the networked visual motion observer is applied to the cameras. In terms of the issue, we have the following theorem.

Theorem 2: Suppose that all the assumptions of Theorem 1 hold and the graph is undirected. Then, if $\varepsilon \geq 1$ we have $T_\varepsilon \leq \tau(c) + \bar{T}_1$ with

$$\bar{T}_1 := \max\left\{ 0, \frac{V(0) - \varepsilon\rho}{\tilde{Q}\beta\lambda_{min2}(L_G)} \left(\frac{1}{k_s} + \frac{\lambda_{max}(L_G)}{k_e} \right) \right\},$$

where $\tilde{Q} = \frac{1}{n}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}} \phi(e^{-\hat{\xi}\theta_{wo_i}} e^{\hat{\xi}\theta_{wo_j}})$, $\lambda_{min2}(M)$ and $\lambda_{max}(M)$ are respectively the second smallest and the largest eigenvalues of matrix M and L_G is the graph Laplacian [7] of graph G. In addition, if $\varepsilon \in [\varepsilon_{ave}, 1)$, then $T_\varepsilon \le \tau(c) + \bar{T}_1 + \max\{T_2, T_3\}$ with

$$\bar{T}_2 := \frac{(1+\beta)(1-\varepsilon)}{k_e \beta},$$

$$\bar{T}_3 := \frac{(1-\varepsilon)\rho}{\tilde{Q}\beta\lambda_{min2}(L_G)}\left(\frac{1}{k_s} + \frac{\lambda_{max}(L_G)}{\epsilon k_e}\right).$$

Proof: Omitted. ■

To make the meaning of the theorem more clear, we show the following corollary, which is immediately proved from Theorem 2.

Corollary 1: Suppose that c, ϵ, k and $\varepsilon(\varepsilon \ge \varepsilon_{ave})$ are fixed (Then, the remaining free parameters are G and k_e). Then, there exist positive scalars c_1, c_2 and c_3 such that

$$T_\varepsilon \le \frac{c_1}{k_e}\left(\frac{\lambda_{max}(L_G)}{\lambda_{min2}(L_G)}\frac{1}{\lambda_{min2}(L_G)} + c_2\right) + c_3.$$

The above results give us helpful insights into the gain selection and network design. Indeed, when the convergence speed is insufficient for a designer, Theorem 2 provides quantitative information on how a redesigned graph and gains speed up the convergence. Meanwhile, Corollary 1 gives qualitative insights. In terms of the network design, Corollary 1 says that if the ratio $\bar{\lambda} = \lambda_{max}(L_G)/\lambda_{min2}(L_G)$ is small, the convergence speed accelerates. The ratio $\bar{\lambda}$ is known to be an important physical quantity reflecting synchronizability of a network [14], [15]. Corollary 1 also says that the algebraic connectivity $\lambda_{min2}(L_G)$ itself should be large to assure high convergence speed even in our problem similarly to the consensus problem on a vector space [7]. Corollary 1 also implies that a large visual feedback gain accelerates the convergence speed, though it might be trivial from the form of the estimation algorithm.

V. Tracking Performance Analysis

In this section, we analyze the tracking performance of the estimates $\{\bar{g}_{io_i}\}_{i\in\mathcal{V}}$ to the average g_i^* for moving targets when the networked visual motion observer is applied to vision cameras under the following assumption.

Assumption 3: (i) $V_{wo_i}^b(t)$ is continuous in t, and $\|\omega_{wo_i}^b(t)\|^2 \le \bar{w}\ \forall i \in \mathcal{V},\ t \ge 0$. (ii) For all time $t \ge 0$, there exists a pair $(i(t), j(t)) \in \mathcal{V}\times\mathcal{V}$ such that $e^{\hat{\xi}\theta_{wo_i}}(t) \ne e^{\hat{\xi}\theta_{wo_j}}(t)$. (iii) $e^{-\hat{\xi}\theta_{wo_j}}(t)e^{\hat{\xi}\theta_{wo_i}}(t) > 0\ \forall i,j \in \mathcal{V}$ and $t \ge 0$.

A. Motion of The Average

In this subsection, we first present a formulation of the average motion other than (8). Note first that $e^{\hat{\xi}\theta^*}$ is continuously differentiable whose proof is shown in [16]. Moreover, since $e^{\hat{\xi}\theta^*}(t) \in SO(3)$ holds for all $t \ge 0$, the derivative $\dot{e}^{\hat{\xi}\theta^*}$ has to satisfy $\dot{e}^{\hat{\xi}\theta^*} \in T_{e^{\hat{\xi}\theta^*}}SO(3) := \{e^{\hat{\xi}\theta^*}X |\ X \in so(3)\}$, where $T_{e^{\hat{\xi}\theta^*}}SO(3)$ is the tangent space of the manifold $SO(3)$ at $e^{\hat{\xi}\theta^*}$. Namely, the trajectory of the Euclidean mean is described by the differential equation

$$\dot{e}^{\hat{\xi}\theta^*} = e^{\hat{\xi}\theta^*}\hat{\omega}^{b,*}$$

with some velocity $\hat{\omega}^{b,*} \in so(3)$.

We next clarify relations between velocities $\omega^{b,*}$ and $\omega_{wo_i}^b$, $i \in \mathcal{V}$. Let us now define $w := (\omega_{wo_i}^b)_{i\in\mathcal{V}}$. Then, we have the following lemma.

Lemma 2: Suppose that $(e^{\hat{\xi}\theta_{wo_i}})_{i\in\mathcal{V}}$ satisfies

$$\left\|e^{\hat{\xi}\theta^*}(t) - S(t)\right\|_F \le \gamma\ \forall t \ge 0 \qquad (16)$$

for some γ. Then, the following inequality holds for all $t \ge 0$.

$$\|\omega^{b,*}(t)\|^2 < \frac{\mu^2(\gamma)}{n}\|w(t)\|^2,\ \mu(\gamma) := \frac{\sqrt{2}}{\sqrt{2}-\gamma} \qquad (17)$$

Proof: See [16]. ■

Note that $\|e^{\hat{\xi}\theta^*}(t) - S(t)\|_F$ is upper bounded by ϕ_m and hence it is estimated by prior information on ϕ_m.

B. Tracking Performance Analysis

We consider the whole networked system consisting of the target object motion (1) and the networked visual motion observer (3), (4) and (9) for all $i \in \mathcal{V}$. Let the collection of body velocities of the target objects $w = (\omega_{wo_i}^b)_{i\in\mathcal{V}}$, be the external input to the systems.

The objective here is to evaluate the distance from the estimates $(\bar{g}_{io_i})_{i\in\mathcal{V}}$ to the average g_i^* in the presence of the disturbance w. Unlike the static objects case, $\rho = \sum_{i\in\mathcal{V}}\phi(e^{-\hat{\xi}\theta_i^*}e^{\hat{\xi}\theta_{io_i}})$ is also time-varying. We thus define

$$\rho' := \sup_{t\ge 0}\rho(t),$$

and redefine the set $\Omega'(\varepsilon)$ by just using ρ' instead of ρ in (12). The parameter ρ' is the supremum of the distance from the estimate to the average when g_{io_i} is correctly estimated and hence it is also an indicator of the best average estimation performance in the absence of communication. Note that the visual motion observer cannot correctly estimate g_{io_i} as long as the object is moving with unknown velocity.

The problem to be considered here is redefined as follows.

Definition 2: The estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ are said to achieve ε-level tracking performance if there exists a finite T s.t.

$$(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}(t))_{i\in\mathcal{V}} \in \Omega'(\varepsilon)\ \forall w \text{ and } t \ge T$$

Then, we have the following theorem.

Theorem 3: Under Assumptions 1 and 3, if (17) holds for some $\gamma > 0$ and $k_e > \mu^2(\gamma)$, then the orientation estimates $(e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})_{i\in\mathcal{V}}$ updated by the networked visual motion observer achieve ε_{track}-level tracking performances with

$$\varepsilon_{track} := 1 + \frac{\mu^2(\gamma)}{k_e - \mu^2(\gamma)} + \frac{\bar{w}^2}{\rho'(k_e - \mu^2(\gamma))}.$$

This theorem implies that ultimate average estimation accuracy improves if the visual feedback gain k_e is large, which is natural from the structure of the estimation scheme.

In summary, to achieve fast convergence and a good tracking performance, we should make the visual feedback gain k_e large (Theorems 2 and 3). However, this results in a poor averaging performance unless we set a sufficiently large mutual gain k_s (Theorem 1). If such a large k_s is acceptable, we have no problem on the gain selection. However, the size of k_s is in general restricted by the communication

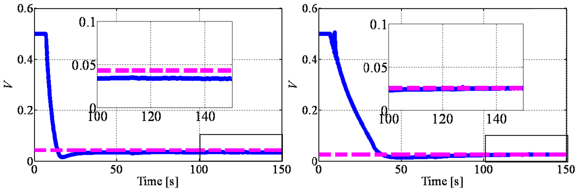

Fig. 3. Time Responses of V (Left: $k_s = 0.2$, Right: $k_s = 10$)

rate due to limitation in standard feedback control theory. Then, a trade-off occurs between averaging and tracking performances. Namely, if we set a large k_e, then a good tracking performance together with high convergence speed is achieved at the cost of a poor averaging performance and vice versa. (See [17] for more details on the gain selection).

VI. Experimental Verification

In this section, we demonstrate the effectiveness of the present scheme by using a visual sensor network testbed with three vision cameras, whose detailed information including visual feedback and communication rates is found in [17]. The cameras have the same orientations and only the positions have the biases $p_{12} = [-0.61\ 0\ 0]^T$[m], $p_{13} = [0\ \ -0.58\ 0]^T$[m]. Let the frame of the camera 1 be the world frame. Here, we employ the communication graph $\mathcal{E} = \{(1,2),(2,1),(2,3),(3,2)\}$ with $d_m = \sqrt{2}$.

Namely, we let the pose estimated by the visual motion observer be g_{wo_i}. In the experimental system, we have $\xi \sin\theta_{wo_1} = [0.115\ \ 0.181\ \ 0.067]^T, \xi \sin\theta_{wo_2} = [-0.149\ 0.156\ 0.026]^T, \xi \sin\theta_{wo_3} = [0.062\ 0.116\ 0.071]^T$, which give the average $\xi \sin\theta^* = [0.009\ \ 0.152\ \ 0.055]^T$. Through trial and error processes, we check that the relative orientation between $e^{\hat{\xi}\theta_{wo_i}}$ and $e^{\hat{\xi}\theta_{wo_j}}$ is upper bounded by $\phi_m = 0.04$, and hence we set $\beta = 0.70$. In addition, let the initial estimates $e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}}(0)$ be $\bar{\xi}\sin\bar{\theta}_{wo_i}(0) = [0.289\ 0.289\ 0.289]^T\ \forall i \in \{1,2,3\}$.

In the experiment, to demonstrate validity of Theorem 1, we choose two different gains $k_e = 0.2$, $k_s = 0.2$ and $k_e = 0.2$, $k_s = 10$, where the first choice does not satisfy $k \le \beta/d_m^2$ ($\varepsilon_{ave} = 1$) and the second one satisfies it ($\varepsilon_{ave} = 0.5987$ for $c = 10^{-4}, \epsilon = 10^{-2}$). Fig. 3 shows the time responses of the function V for $k_s = 0.2$ and $k_s = 10$. In these figures, the magenta lines indicate the value $\varepsilon_{ave}\rho$. We see from the bottom right figures that the responses of V are eventually lower than the corresponding lines as proved in Theorem 1. In addition, we also see that the larger mutual feedback gain achieves a more accurate average estimation than the smaller one. More detailed analysis on the experiments is shown in [17]. The movie of the experiment is available at `http://www.fl.ctrl.titech.ac.jp/researches/movie_new/movie7/vsn_ce.wmv`.

Though a quantitative evaluation of tracking performance cannot be addressed due to the difficulties in computing the average motion, the latter half of the above movie sufficiently supports validity of our claim that a large visual feedback gain results in a good tracking performance. Verifications through simulation are also addressed in [16].

VII. Conclusions

In this paper, we have addressed cooperative estimation of 3D target object motion via networked visual motion observers. After introducing the networked visual motion observers, we have clarified the ultimate averaging accuracy, a relation between the convergence speed and design parameters in the algorithm and tracking performance for moving target objects. 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] Y. Ma, S. Soatto, J. Kosecka and S. S. Sastry, "An Invitation to 3-D Vision: From Images to Geometric Models," Springer, 2004.

[4] I. J. Ndiour and P. A. Vela, "A Local Extended Kalman Filter for Visual Tracking (I)," Proc. of the 49th IEEE Conference on Decision and Control, pp. 2498–2504, 2010.

[5] 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.

[6] T. Hatanaka and M. Fujita "Passivity-based Visual Motion Observer: From Theory to Distributed Algorithms," Proc. of the IEEE 2010 Multi-conference on Systems and Control, (Tutorial Session Paper), pp. 1210–1221, 2010.

[7] 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.

[8] 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.

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

[10] A. Edelmayer, M. Miranda and V. Nebehaj, "Cooperative Federated Filtering Approach for Enhanced Position Estimation and Sensor Fault Tolerance in Ad-hoc Vehicle Networks," IET Intelligent Transport Systems, Vol. 4, No. 1, pp. 82–92, 2010.

[11] 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.

[12] T. Hatanaka and M. Fujita "Passivity-based Cooperative Estimation of 3D Target Motion for Visual Sensor Networks: Analysis on Averaging Performance," Proc. of 2011 American Control Conference, pp. 3399–3404, 2011.

[13] 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.

[14] L. Donetti, P. I. Hurtado, and M. A. Munoz, "Entangled Networks, Synchronization, and Optimal Network Topology," Physical Review Letters, Vol. 95, No. 18, 2005.

[15] T. Li, M. Fu, L. Xie and J. Zhang, "Distributed Consensus With Limited Communication Data Rate," IEEE Trans. on Automatic Control, Vol. 56, No. 2, pp. 279–292, 2011.

[16] T. Hatanaka and M. Fujita, "Cooperative Estimation of 3D Target Object Motion via Networked Visual Motion Observers," IEEE Trans. on Automatic Control, submitted, 2011 (available at arXiv:1107.5108).

[17] T. Hatanaka, T. Nishi and M. Fujita, "Passivity-based Cooperative Estimation Algorithm for Networked Visual Motion Observers," SICE Annual Conference 2011, 2011 (to appear) http://www.fl.ctrl.titech.ac.jp/paper/2011/HNF_sice11.pdf.