Vision-based Cooperative Estimation via Multi-agent Optimization

Takeshi Hatanaka, Masayuki Fujita and Francesco Bullo

***Abstract*— In this paper, we investigate a cooperative estimation problem for visual sensor networks based on multi-agent optimization techniques. A passivity-based visual motion observer is employed as a tool to meet the objective. We first give an interpretation of the visual motion observer from the viewpoint of optimization and present new inputs motivated by the optimization techniques on manifolds. Based on the investigations, we formulate a novel cooperative estimation problem to be tackled. Then, a cooperative estimation algorithm is presented based on multi-agent optimization techniques. Finally, the effectiveness of the present algorithm is demonstrated through experiments.**

I. Introduction

A visual sensor network [1]–[3] 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 such explicit data but combining image processing techniques or human operators with the measurement gives information on what happens, what a target is, where it is and where it bears. 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 visual servoing [4]–[8]. The authors also presented dynamic visual feedback control schemes for 3D target tracking based on passivity in [7], where a vision-based observer called visual motion observer plays a central role to estimate the target's pose. However, in visual sensor networks, it is expected not only to give an estimate but also to cooperate with each other vision sensor, which brings us new theoretical challenges. The main advantages of cooperation are: (i) accuracy of estimates by integrating richer information than the case of a single sensor, (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.

Cooperative estimation for sensor networks has been tackled in recent years [9], [10]. [10] presents distributed Kalman filters based on the consensus algorithm [11] and exemplifies the fact that averaging the estimates among the neighbors achieves more accurate estimation than averaging sensed data as in [9]. Unfortunately, the algorithm is not applicable to our problem since the object's pose takes values in a non-Eucledean space. Meanwhile, [2] presents a distributed version of the computation algorithm of an average on Special Orthogonal Group called Karcher mean [12]. However, this work focuses on the averaging by assuming that the target's orientation is obtained *a priori* and does not mention estimation from image data.

In this paper, we formulate a novel vision-based cooperative estimation problem for visual sensor networks and present an algorithm based on multi-agent optimization techniques [13]. For this purpose, after introducing the problem setting and the visual motion observer [7], we first reconsider the processes in the observer from the viewpoint of optimization. Then, we present some novel inputs motivated by the techniques in optimization and prove correctness of the estimate based on passivity. We moreover show that using passivity allows us to evaluate the error from the actual object's pose even if the object is moving, while optimization handles only static problems. Following the above investigations, we next formulate a cooperative estimation problem, where a minimization problem of an aggregate objective function is presented. We then present a cooperative estimation algorithm which embody both of the consensus and gradient descent algorithm. The gradient descent term is introduced to make the estimate track to a moving object's pose while the consensus term is to lead the estimate to a value close to an average of vision sensors' estimates if the object is static. We moreover gives an upper bound of the error between the estimates and their average for a static object based on the results in the multi-agent optimization [13]. Finally, it is demonstrated through experiments that the present algorithm is effective in estimation for both static and moving target object.

II. Notations and Basis of Optimization on $SO(3)$

For a matrix $M \in \mathbb{R}^{3\times 3}$, M^T denotes the transpose of M, $\mathrm{tr}(M)$ the trace of M, and $\det(M)$ the determinant of M. The operators $\mathrm{sym}(\cdot)$ and $\mathrm{sk}(\cdot)$ are respectively defined as $\mathrm{sym}(M) = \frac{1}{2}(M + M^T)$ and $\mathrm{sk}(M) = \frac{1}{2}(M - M^T)$.

Let us first consider the Lie group $SO(3)$ defined as below.

$$SO(3) = \{R \in \mathbb{R}^{3\times 3} |\ R^T R = I_3,\ \det(R) = +1\}$$

We describe the vector space consisting of all skew symmetric matrices in $\mathbb{R}^{3\times 3}$ by $so(3)$. The notation $\wedge : \mathbb{R}^3 \to so(3)$ is a skew symmetric operator satisfying $\hat{x}y = x \times y$ for any vectors x and y with cross product $\times$. $\vee : so(3) \to \mathbb{R}^3$ is the inverse operator of $\wedge$. The exponential map from $so(3)$ to $SO(3)$ is denoted by $\exp(\cdot)$. We usually use $e^{\hat{\omega}} \in SO(3)$ to denote $\exp(\hat{\omega})$. The inverse map of the exponential map is

Takeshi Hatanaka and Masayuki Fujita are with the Department of Mechanical and Control Engineering, Tokyo Institute of Technology, Tokyo 152-8552, JAPAN, Francesco Bullo is with Department of Mechanical Engineering at the University of California at Santa Barbara

denoted by $\log(\cdot)$. For more details on the notations, please refer to [14].

In this paper, we consider the following two metrics on $SO(3)$ between any two elements Q and R of $SO(3)$.

$$\phi_Q^{\mathrm{F}}(R) = \frac{1}{2}\|R - Q\|_F^2 = \mathrm{tr}(I - Q^T R) \tag{1}$$

$$\phi_Q^{\mathrm{R}}(R) = \frac{1}{2}\|\log(Q^T R)\|_F^2 \tag{2}$$

where $\|\cdot\|_F$ represents the Frobenius norm. The function $\phi_Q^{\mathrm{F}}(R)$ and $\phi_Q^{\mathrm{R}}(R)$ are square of so-called Euclidean distance and Riemannian distance [15] respectively. It is known that the gradients [16] of $\phi_Q^{\mathrm{F}}(R)$ and $\phi_Q^{\mathrm{R}}(R)$ at $R \in SO(3)$ are given as follows [2], [12], [15].

$$\mathrm{grad}_R \phi_Q^{\mathrm{F}} = -R\mathrm{sk}(R^T Q), \tag{3}$$

$$\mathrm{grad}_R \phi_Q^{\mathrm{R}} = -R\log(R^T Q). \tag{4}$$

The Newton vector with respect to the function $\phi_Q^{\mathrm{F}}(R)$ at $R \in SO(3)$ is η_R satisfying

$$\mathrm{Hess}_R\ \phi_Q^{\mathrm{F}}\eta_R = -\mathrm{grad}_R\ \phi_Q^{\mathrm{F}}(R), \tag{5}$$

where $\mathrm{Hess}_R\ \phi_Q^{\mathrm{F}}\eta_R$ is the Hessian of ϕ_Q^{F} at R [16]. The solution to (5) is given by the closed form

$$\begin{aligned} \eta_R &= R\gamma_{\mathrm{nv}}(R^T Q)^{\wedge}, \qquad (6)\\ \gamma_{\mathrm{nv}}(M) &= \bar{\gamma}_{\mathrm{nv}}^{-1}(M)\mathrm{sk}(M)^{\vee},\\ \bar{\gamma}_{\mathrm{nv}}(M) &= \frac{1}{2}\left(\mathrm{tr}(\mathrm{sym}(M))I_3 - \mathrm{sym}(M)\right). \end{aligned}$$

We next consider the special Euclidean space $SE(3) := \mathbb{R}^3 \times SO(3)$. If we use the homogeneous representation, each element $g = (p, R) \in SE(3)$ is described as

$$g = \begin{bmatrix} R & p \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}. \tag{7}$$

Analogous to the definition of $so(3)$, we define $se(3) = \mathbb{R}^3 \times so(3)$. Then, in homogeneous coordinates, an element $\hat{\xi} \in se(3)$ is described as

$$\hat{\xi} = \begin{bmatrix} \hat{\omega} & v \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 4}.$$

In this paper, we also use the following metric on $SE(3)$ between two elements $g_1 = (p_1, R_1)$ and $g_2 = (p_2, R_2)$.

$$\psi_{g_1}(g_2) = \varphi_{p_1}(p_2) + \phi_{R_1}^{\mathrm{F}}(R_2), \tag{8}$$

where $\varphi_q(p) = \frac{1}{2}\|p - q\|^2$.

III. Problem Setting

Throughout this paper, we consider the situation where n vision cameras see different target objects (Fig. 1). Suppose that each vision camera $i \in \mathcal{V} := \{1, \cdots, n\}$ has communication and computation capability. The problem is motivated by some scenarios such as estimation of group behaviors, estimation under uncertainties including noises, incomplete localization and parametric uncertainties of vision cameras. With such uncertainties, the visual measurement would be contaminated by them and the object's pose consistent with the measurement would differ among sensors even if the actual target is single. Under the situation, a way to gain accurate estimate of the object is averaging the individual objects' poses.

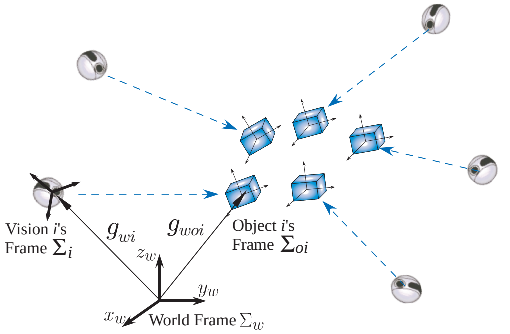

Fig. 1. Visual Sensor Network

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 which vision camera i sees, respectively. Then, the pose of vision camera i and object o_i 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 relative pose from Σ_i to Σ_{o_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 $V_{wo_i}^b = (v_{wo_i}, \omega_{wo_i})$ or

$$\hat{V}_{wo_i}^b = g_{wo_i}^{-1}\dot{g}_{wo_i} = \begin{bmatrix} \hat{\omega}_{wo_i} & v_{wo_i} \\ 0 & 0 \end{bmatrix} \in \mathcal{R}^{4\times 4}, \tag{9}$$

where v_{wo_i} and ω_{wo_i} represent the linear velocity of the origin and the angular velocity from Σ_w to Σ_c, respectively [14]. Similarly, vision camera i's body velocity relative to Σ_w will be denoted as $V_{wi}^b = (v_{wi}, \omega_{wi})$.

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

$$V_{io_i}^b = -\mathrm{Ad}_{(g_{io_i}^{-1})}V_{wi}^b + V_{wo_i}^b. \tag{10}$$

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 [14].

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. With a slight abuse of notation, as $[p_{il}^T\ 1]^T$ and $[p_{o_il}^T\ 1]^T$, we have $p_{il} = g_{io_i}p_{o_il}$,.

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

[14] that the perspective projection of the l-th feature point onto the image plane gives us the image plane coordinate $f_{il} \in \mathcal{R}^2$ as

$$f_{il} \quad = \quad \frac{\lambda_i}{z_{il}} \begin{bmatrix} x_{il} \\ y_{il} \end{bmatrix}, \tag{11}$$

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 $p_i := [p_{i1}^T \ \cdots \ p_{im}^T]^T \in \mathcal{R}^{3m}$.

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 $(j, i) \in \mathcal{E}$. In addition, we define the neighbor set $\mathcal{N}_i := \{j \in \mathcal{V} | \ (j, i) \in \mathcal{E}\}$.

IV. 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].

A. Estimation Error System

The visual motion observer has the same structure as Luenberger observer. We thus first prepare the model of the actual rigid body motion relative to the vision camera. Using the relative rigid body motion (10), we choose estimates $\bar{g}_{io_i} = (\bar{p}_{io_i}, e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}})$ and $\bar{V}_{io_i}^b$ of the relative rigid body motion and velocity respectively as

$$\bar{V}_{io_i}^b \quad = \quad -\mathrm{Ad}_{(\bar{g}_{io_i}^{-1})} V_{wi}^b + u_{ei}. \tag{12}$$

The new input $u_{ei} = [v_{uei}^T \ \omega_{uei}^T]^T$ is to be determined in order 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 $\bar{f}_i$ $(i = 1, \ldots, m)$ is also computed similarly to (III-B) and (11). In the following, we use the notation $\bar{p}_{il} := [\bar{x}_{il} \ \bar{y}_{il} \ \bar{z}_{il}]^T, \ l \in [1, m]$.

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 $e_R(e^{\hat{\xi}\theta}) := \mathrm{sk}(e^{\hat{\xi}\theta})^\vee$, the vector representation of the estimation error is also given by

$$e_{ei} := E_R(g_{ei}), \ E_R(g_{ei}) := \begin{bmatrix} p_{ei}^T & e_R^T(e^{\hat{\xi}\theta_{ei}}) \end{bmatrix}^T. \tag{13}$$

Note that $e_{ei} = 0$ iff $p_{ei} = 0$ and $e^{\hat{\xi}\theta_{ei}} = I_3$.

Let us now derive a relation between the actual and estimated visual measurements. If we define the visual error as $f_{ei} := f_i(g_{io_i}) - \bar{f}_i(\bar{g}_{io_i})$, then the estimation error vector e_{ei} can be reconstructed from the visual error by

$$e_{ei} = J_i^\dagger(\bar{g}_{io_i}) f_{ei}, \tag{14}$$

[7], where † denotes the pseudo-inverse and $J_i(\bar{g}_{io_i}) : SE(3) \to \mathcal{R}^{2m \times 6}$ is defined as

$$\begin{aligned} J_i(\bar{g}_{io_i}) \quad &:= \quad \begin{bmatrix} J_{i1}^T(\bar{g}_{io_i}) & J_{i2}^T(\bar{g}_{io_i}) & \cdots & J_{im}^T(\bar{g}_{io_i}) \end{bmatrix}^T, \\ J_{il}(\bar{g}_{io_i}) \quad &:= \quad \tilde{J}_{il}(\bar{g}_{io_i}) e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}} \begin{bmatrix} I & -\hat{p}_{o_i l} \end{bmatrix}, \\ \tilde{J}_{il}(\bar{g}_{io_i}) \quad &:= \quad \begin{bmatrix} \frac{\lambda}{\bar{z}_{il}} & 0 & -\frac{\lambda \bar{x}_{il}}{\bar{z}_{il}^2} \\ 0 & \frac{\lambda}{\bar{z}_{il}} & -\frac{\lambda \bar{y}_{il}}{\bar{z}_{il}^2} \end{bmatrix}. \end{aligned}$$

We assume that the matrix $J_i(\bar{g}_{io_i})$ is full column rank for all $\bar{g}_{io_i} \in SE(3)$. It is known that if $m \geq 4$ the image Jacobian has the full column rank.

Differentiating $g_{ei} = \bar{g}_{io_i}^{-1} g_{io_i}$ with respect to time and using (10) and (12), we obtain the estimation error system

$$V_{ei}^b \quad = \quad -\mathrm{Ad}_{(g_{ei}^{-1})} u_{ei} + V_{wo_i}^b. \tag{15}$$

B. Stability Analysis

In this subsection, we design the visual motion observer and analyze stability of the closed-loop system. For this purpose, we first give a remarkable fact.

Fact 1: [7] If $V_{wo_i}^b = 0$, then the following inequality holds for the estimation error system (15).

$$\int_0^T u_{ei}^T(-e_{ei}) dt \geq -\gamma_i, \tag{16}$$

where γ_i is a positive scalar.
Let us take u_{ei} as the input and e_{ei} as the output of (15). Then, Fact 1 implies that the estimation error system (15) is *passive* from the input u_{ei} to the output $-e_{ei}$.

Based on the above passivity property, we consider the following input

$$u_{ei} = -k_{ei}(-e_{ei}) = k_{ei} e_{ei}, \ k_{ei} > 0. \tag{17}$$

Then, from passivity-based control theory, we can prove the asymptotic stability of the equilibrium point $e_{ei} = 0$ for the closed-loop system (15) and (17). This implies that the visual motion observer leads the estimate $\bar{g}_{io_i}$ to the actual relative pose of the static object g_{io_i} asymptotically as long as the initial estimation error is small enough.

V. Reconsideration of Visual Motion Observer

In this section, we reconsider the update procedure of the estimate $\bar{g}_{io}$ in the visual motion observer from the viewpoint of optimization on $SE(3)$.

In case of $V_{wo_i}^b = V_{wi}^b = 0$, we see from (12) and (17) that the estimate $\bar{g}_{io_i}$ is updated according to

$$\begin{aligned} \dot{\bar{g}}_{io_i} \quad &= \quad \bar{g}_{io_i} (k_{ei} e_{ei})^\wedge \\ &= \quad k_{ei} \begin{bmatrix} e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}} \mathrm{sk}(e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}} e^{\hat{\xi}\theta_{io_i}}) & p_{io_i} - \bar{p}_{io_i} \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{18}$$

From (3), (18) is rewritten as

$$\dot{\bar{g}}_{io} = k_{ei} \begin{bmatrix} -\mathrm{grad}_{e^{\hat{\bar{\xi}}\bar{\theta}_{io_i}}} \phi^{\mathrm{F}}_{e^{\hat{\xi}\theta_{io_i}}} & -\mathrm{grad}_{\bar{p}_{io_i}} \varphi_{p_{io_i}} \\ 0 & 0 \end{bmatrix}. \tag{19}$$

(19) indicates that the update process of the estimate is interpreted as a process to solve the optimization problem

$$\min_{\bar{g}\in SE(3)} \psi_{g_{io_i}}(\bar{g}). \tag{20}$$

From the fact, we will use the function ψ as an individual objective function to be minimized in cooperative estimation for visual sensor networks. Before mentioning it, we give some extensions of the results in [7] from the viewpoint of optimization theory on manifolds [16] in this section.

A. Riemannian Metric

In this subsection, we use the Riemannian metric $\phi^{\rm R}$ instead of $\phi^{\rm F}$ in (20), which gives the update procedure

$$u_{ei} = u^{\rm R}(e_{ei}) = k_{ei}\gamma(e_{ei}),\ |\theta_{ei}| < \pi/2. \tag{21}$$
$$\gamma([x_1^T\ x_2^T]^T) := \begin{bmatrix} x_1 \\ \gamma_R(x_2) \end{bmatrix},\ \gamma_R(x) := \bar{\gamma}_R(x)x$$
$$\bar{\gamma}_R(x) := \sin^{-1}(\|x\|)/\|x\|$$

The restriction $|\theta_{ei}| < \pi/2$ is imposed to guarantee well-definedness of the function $\bar{\gamma}_R(\cdot)$.

In terms of (21) we immediately obtain the following proposition.

Proposition 1: Suppose that $V^b_{wo_i} = 0$. Then, for the closed-loop system (12) with the input (21), the origin $e_{ei} = 0$ is an asymptotically stable equilibrium point.

In addition, we have the following result for a moving object.

Proposition 2: Consider the system (12) and (21) with input $V^b_{wo_i}$ and output e_{ei}. Then, if

$$k_{ei} - \frac{1}{2\delta^2} - \frac{1}{2} > 0 \tag{22}$$

holds for some $\delta > 0$ and $|\theta_{ei}| < \pi/2$ is satisfied for all time, the L_2-gain of the system is less than δ.

B. Newton Method

To accelerate convergence in optimization, we have an option to use Newton method. Namely, in this subsection, we use instead of (19) the update rule

$$u_{ei} = u^{\rm nv}(e_{ei}) := \begin{bmatrix} I & 0 \\ 0 & \bar{\gamma}^{-1}_{\rm nv}(e^{\hat{\xi}\theta_{ei}}) \end{bmatrix} e_{ei},\ |\theta_{ei}| < \pi/2. \tag{23}$$

In terms of (23), we have the following proposition.

Proposition 3: Suppose that $V^b_{wo_i} = 0$. Then, for the closed-loop system (12) with the input (23), the origin $e_{ei} = 0$ is an asymptotically stable equilibrium point.

We also get the following proposition for a moving object.

Proposition 4: Consider the system (12) and (23) with input $V^b_{wo_i}$ and output e_{ei}. Then, if

$$k_{ei} - \frac{1}{2\delta^2} - \frac{1}{2} > 0 \tag{24}$$

holds for some $\delta > 0$ and $|\theta_{ei}| < \pi/2$ is satisfied for all time, the L_2-gain of the system is less than δ.

Propositions 1 and 3 prove only convergence and a similar statement might be obtained by optimization theory once e_{ei} is reconstructed from the image data though the reconstruction is also a part of the visual motion observer. However, optimization theory handles only static problems and does not provide any answer to the estimation when the sensor and target object are moving as in Propositions 2 and 4.

VI. Cooperative Estimation Algorithm

A. Problem Reformulation

In this section, we consider a visual sensor network consisting of multiple vision sensors assuming that each vision sensor i knows g_{wi}.

We first formulate the cooperative estimation problem to be considered in this paper. It is expected for visual sensor networks to meet the following requirements simultaneously.

- (Averaging) The estimates take values close to an average of $\{g_{wo_i}\}_{i\in\mathcal{V}}$ for a static object.
- (Tracking) The estimates track the actual object's pose g_{wo_i} for a moving object.

We first define the individual objective function to be minimized by each vision sensor i as $\psi_{g_{wo_i}}$. Then, we formulate a cooperative estimation problem as a minimization of the aggregate objective function

$$\min_{\bar{g}=(\bar{p},e^{\hat{\bar{\xi}}\bar{\theta}})\in SE(3)} \Psi(\bar{g}),\ \Psi(\bar{g}) := \frac{1}{n}\sum_{i=1}^{n} \psi_{g_{wo_i}}(\bar{g}). \tag{25}$$

It should be now noted that in our setting each vision sensor does not know neighbors' objective function $\psi_{g_{wo_j}}$ since it contains g_{wo_j} to be estimated.

The problem (25) is divided into independent problems

$$\min_{\bar{p}\in\mathbb{R}^3} \frac{1}{n}\sum_{i=1}^{n} \varphi_{p_{wo_i}}(\bar{p}), \tag{26}$$

$$\min_{e^{\hat{\bar{\xi}}\bar{\theta}}\in SO(3)} \Phi(e^{\hat{\bar{\xi}}\bar{\theta}}),\ \Phi(e^{\hat{\bar{\xi}}\bar{\theta}}) := \frac{1}{n}\sum_{i=1}^{n} \phi^{\rm F}_{e^{\hat{\xi}\theta_{wo_i}}}(e^{\hat{\bar{\xi}}\bar{\theta}}). \tag{27}$$

The solution to (27) is called Euclidean mean [15]. Given orientations $e^{\hat{\xi}\theta_{wo_i}},\ i\in\mathcal{V}$, the mean R^* is given by

$$e^{\hat{\xi}\theta^*} = {\rm Proj}(\bar{R}),\ \bar{R} = \frac{1}{n}\sum_{i\in\mathcal{V}} e^{\hat{\xi}\theta_{wo_i}} \tag{28}$$

[15], where ${\rm Proj}(M)$ gives the orthogonal projection of M onto $SO(3)$ and is computed effectively by using the singular value or polar decompositions of M [17]. Note that just computing the Euclidean mean for a static object is not so difficult even in a distributed fashion. Indeed, $\bar{R}$ is computed by using the consensus algorithm [11] under appropriate assumptions on the graph. However, such a scheme works only for a static object.

B. Multi-agent Optimization

In this paper, we present an algorithm, which embodies not only the consensus but also the the gradient descent algorithm with $\psi_{g_{wo_i}}$ in order to achieve tracking. Recently, such an update procedure is presented in [13] in order to solve the multi-agent optimization problem

$$\min_{x\in\mathbb{R}^n} F(x) := \sum_{i=1}^{N} F_i(x)\ (F_i : \text{convex}), \tag{29}$$

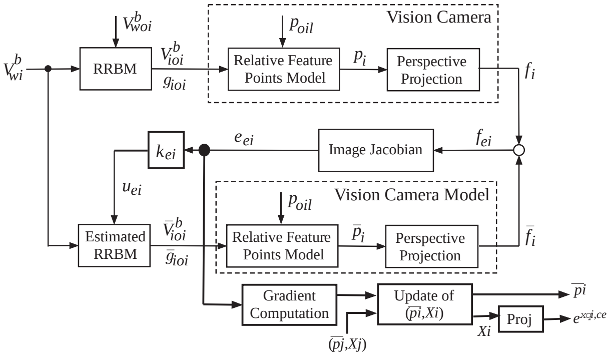

Fig. 2. Block Diagram of Cooperative Estimation Algorithm

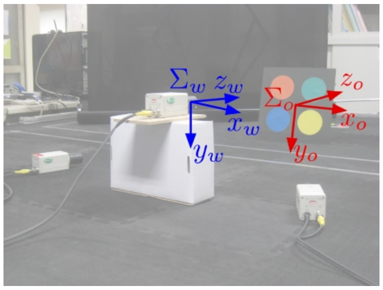

Fig. 3. Overview of Experimental Environment

under the situation where each agent i does not know $F_j,\ j \neq i$ but only F_i. There, the following update rule of the estimate of solution x_i is presented, which consists of the consensus and gradient descent algorithm of the individual objective function F_i.

$$x_i[k+1] = -\alpha_i \mathrm{grad}_{x_i[k]} F_i + \sum_{j \in \mathcal{V}} a_{ij} x_j[k], \tag{30}$$

where $a_{ij} = a_{ji},\ a_{ij} = 0$ if $(i,j) \notin \mathcal{E}$ and $\sum_{j \in \mathcal{V}} a_{ij} = 1\ \forall i \in \mathcal{V}$. We remark that (30) is a simplified version and more general one with a variable gain for random graphs is presented in [13]. [13] also derives a bound $\epsilon_{k,i}$ such that

$$F(y_i[k]) \leq F(x^*) + \epsilon_{k,i},\ y_i[k] = \frac{1}{k}\sum_{h=0}^{k-1} x_i[k],$$

where x^* is the actual optimal solution to (29). If we can present such an algorithm for our problem, it is possible to get an approximate value of the Euclidean mean for a static object. Hereafter, we mainly focus on (27) since the minimization (26) is just a special case of the problem [13].

C. Cooperative Estimation Algorithm

In this subsection, we choose $\mathbb{R}^{3\times 3}$ as a vector space to execute averaging. We consider the scalar field $\tilde{\phi}_Q : \mathbb{R}^{3\times 3} \to \mathbb{R}$ such that $\tilde{\phi}_Q(R) = \frac{1}{2}\|R - Q\|_F^2$. Then, the solution to

$$\min_{R \in \mathbb{R}^{3\times 3}} \tilde{\Phi}(R),\ \tilde{\Phi}(R) := \frac{1}{n}\sum_{i=1}^{n} \tilde{\phi}_{e^{\hat{\xi}\theta_{wo_i}}}(R) \tag{31}$$

is given by $\bar{R} = \frac{1}{n}\sum_{i=1}^{n} e^{\hat{\xi}\theta_{wo_i}}$.

Let us now present the update rule of the cooperative estimate $e^{\hat{\bar{\xi}}\bar{\theta}_{i,ce}}[k]$ by introducing a fictitious variable $X_i[k]$.

$$X_i[k+1] = -\alpha_i \mathrm{grad}_{X_i[k]} \tilde{\phi}_{e^{\hat{\xi}\theta_{wo_i}}} + \sum_{j \in \mathcal{V}} a_{ij} X_j[k]. \tag{32}$$

$$e^{\hat{\bar{\xi}}\bar{\theta}_{i,ce}}[k+1] = \mathrm{Proj}(X_i[k+1]) \tag{33}$$

Let us now show how to implement (32) by using the visual motion observer. It is sufficient to show that

$$\mathrm{grad}_X \tilde{\phi}_{e^{\hat{\xi}\theta_{wo_i}}} = X - e^{\hat{\xi}\theta_{wo_i}} \tag{34}$$

is provided by the observer. Once e_{ei} is reconstructed from (14) in the visual motion observer, we get $e^{\hat{\xi}\theta_{ei}} = e^{-\hat{\bar{\xi}}\bar{\theta}_{io_i}} e^{\hat{\xi}\theta_{io_i}} = e^{-\hat{\bar{\xi}}\bar{\theta}_{wo_i}} e^{\hat{\xi}\theta_{wo_i}}$ by using the function $\exp \circ \gamma_R$. Now, we compute M_i such that $M_i^T M_i = 2I - 2\mathrm{sym}(e^{\hat{\xi}\theta_{ei}})$ for the symmetric matrix $2I - 2\mathrm{sym}(e^{\hat{\xi}\theta_{ei}})$, which gives $e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} - e^{\hat{\xi}\theta_{wo_i}}$ since

$$2I - 2\mathrm{sym}(e^{\hat{\xi}\theta_{ei}}) = (e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} - e^{\hat{\xi}\theta_{wo_i}})^T (e^{\hat{\bar{\xi}}\bar{\theta}_{wo_i}} - e^{\hat{\xi}\theta_{wo_i}}).$$

We thus use $X_i[k] + M_i[k] - e^{\hat{\xi}\theta_{wo_i}}$ as the gradient (34).

The total algorithm is shown in Fig. 2, where $\bar{g}_{i,ce}[k]$ is an eventual estimate of sensor i. It should be noted that the estimate $\bar{g}_{io}$ in the visual motion observer is not replaced by $g_{wi}^{-1}\bar{g}_{i,ce}[k]$. This is because the task imposed on the observer is to provide $\mathrm{grad}_{X_i[k]} \tilde{\phi}_{e^{\hat{\xi}\theta_{wo_i}}}$ as accurately as possible and it is achieved when $|\theta_{ei}|$ is sufficiently small.

In the following, we see that an upperbound on the error from the actual Euclidean mean is given by the present algorithm if the object is static. An additional computation of $Y_i[k] = \frac{1}{k}\sum_{h=0}^{k-1} X_i[h]$ immediately gives $\epsilon_{k,i}$ such that $\tilde{\Phi}(X_i[k]) \leq \epsilon_{k,i} + \tilde{\Phi}(\bar{R})$ from [13]. More importantly, [13] provides $\epsilon_i = \lim_{k\to\infty} \epsilon_{k,i}$ and hence, after $X_i[k]$ converges to a value X_i,

$$\tilde{\Phi}(X_i[k]) \leq \epsilon_i + \tilde{\Phi}(\bar{R})$$

since otherwise the statement in [13] does not hold. We thus get an upper bound ϵ_i after convergence without any additional computation. Note however that closeness of $X_i[k]$ to $\bar{R}$ does not imply that of $e^{\hat{\bar{\xi}}\bar{\theta}_{i,ce}}[k]$ to the solution to the original problem (27) in terms of the metric Φ. We thus finally give the following proposition.

Proposition 5: Suppose that $\tilde{\Phi}(X) \leq \epsilon + \tilde{\Phi}(\bar{R})$. Then,

$$\Phi(\mathrm{Proj}(X)) \leq \Phi(e^{\hat{\xi}\theta^*}) + \frac{1}{2}(\sqrt{2\epsilon} + \|\Sigma_X - I\|)^2, \tag{35}$$

holds, where Σ_X is given by the singular value decomposition of X as $X = U_X \Sigma_X V_X^T$.

(35) evaluates the closeness of $e^{\hat{\bar{\xi}}\bar{\theta}_{i,ce}}[k]$ to the solution to the original problem (27). From the fact that $\|\Sigma_X - I\|_F = 0$ if $X \in SO(3)$, (35) implies that the approximate solution degrades as $X_i[k]$ becomes far from $SO(3)$.

VII. Verification through Experiments

We finally demonstrate the effectiveness of the present algorithm through experiments by using three CCD cameras KMT1607 (Komoto Corp.) with lens LTV2Z3314CS-IR (Raymax Corp.). As an object, we prepare a board with

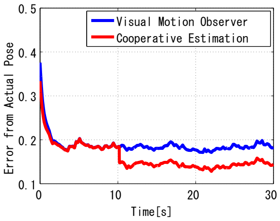

Fig. 4. Error between Estimates and Actual Pose

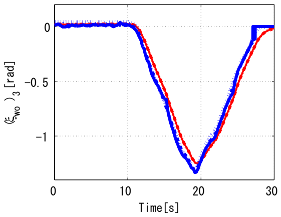

Fig. 5. Time Response of Third Element of $\xi\theta$

four colored feature points attached to a mobile robot e-nuvo WHEEL (ZMP Corp.). A PC can send the velocity command to the robot through a wireless communication device Wiport (LANTRONIX) and it moves according to the command. The image data is sent to a frame grabber board Piccolo Diligent (Euresys Corp.) attached to a PC and the feature points are extracted by an image processing software HALCON (Linx). The present algorithm is written in MATLAB and SIMULINK, and is implemented on a digital signal processor DS1104 (dSPACE Inc.) using the Real-Time Workshop. The frame rate provided by the camera are 30 [fps].

We perform two different experiments. In both experiments, cameras are set so that $p_{w1} = [-0.35\ \ 0.18\ \ 0]^T, \xi\theta_{w1} = [0\ \ 0.37\ \ 0]^T$, $p_{w2} = 0, \xi\theta_{w2} = 0$ $p_{w3} = [0.37\ \ 0.18\ \ 0]^T, \xi\theta_{w3} = [0\ \ -0.37\ \ 0]^T$. Let the initial values be equal to the initial estimates in the visual motion observer as $X_i[0] = e^{[0\ 0\ \hat{\pi}/3]^T}\ \forall i \in \{1, 2, 3\}$. In addition, we use the communication graph with the edge set $\mathcal{E} = \{(1, 2), (2, 1), (2, 3), (3, 2)\}$.

In the first experiment, we run the algorithm for a static object with $\xi\theta_{wo} = [0\ \ 0\ \ 0]^T$. Fig. 4 shows the time responses of the sum of the errors between the estimates and the actual pose measured by the function ϕ, where we start communication at 10[s]. We see from the figure that the function decreases by about $20 - 25$[%] after starting communication, which shows that the estimation accuracy improves due to the averaging term in (32).

In the second experiment, we rotate the object so that the third element of $\xi\theta_{wo}$ decreases over the time interval about $[10, 18]$ and increases over about $[18, 27]$. Fig. 5 shows the time responses of the estimates $(\xi\theta_{wo})_3$ by the cooperative estimation algorithm (red curves) and the visual motion observer (blue curves). The solid curves show the estimates of camera 1, dotted ones those of camera 2 and dash-dotted ones those of camera 3. We see from the figure that tracking to the moving object is also achieved by the present algorithm due to the gradient descent term in (32). The tracking might be seen more apparently by a movie, which is downloaded at `http://www.fl.ctrl.titech.ac.jp/researches/movie_new/movie1/VSN.wmv`. All of these results show the effectiveness of the present algorithm for both a static and moving object.

VIII. Conclusion

In this paper, we have investigated a cooperative estimation problem for visual sensor networks based on multi-agent optimization techniques. A passivity-based visual motion observer has been employed as a tool to meet the objective. We first have given an interpretation of the visual motion observer from the viewpoint of optimization and present new inputs to the observers. Based on the investigations, we have formulated a novel cooperative estimation problem and a cooperative estimation algorithm has been presented based on multi-agent optimization techniques. Finally, the effectiveness of the present algorithm has been demonstrated through experiments.

References

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

[2] R. Tron, R. Vidal and A. Terzis, “Distributed Pose Averaging in Camera Sensor Networks via Consensus on SE(3),” International Conference on Distributed Smart Cameras, 2008.

[3] 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.

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

[5] G. Hu, W. MacKunis, N. Gans, W. E. Dixonm, J. Chen, A. Behal and D. Dawson, “Homography-based Visual Servo Control with Imperfect Camera Calibration,” IEEE Trans. on Automatic Control, Vol. 54, No. 6, pp. 1318–1324, 2009.

[6] T. Ding, M. Sznaier and O. Camps, “Receding Horizon Rank Minimization Based Estimation with Applications to Visual Tracking,” Proc. of the 47th IEEE Conference on Decision and Control, pp. 3446–3451, 2008.

[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] H. Kawai, T. Murao and M. Fujita, “Visual Motion Observer-based Pose Control with Panoramic Camera via Passivity Approach,” Proc. of the 2010 American Control Conference, to appear, 2010.

[9] R. Olfati-Saber, “Distributed Kalman Filter with Embedded Consensus Filters,” Proc. of the 44th IEEE Conference on Decision and Control and 2005 European Control Conference, pp.8179- 8184, 2005.

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

[11] 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.

[12] J. H. Manton, “A Globally Convergent Numerical Algorithm for Computing the Centre of Mass on Compact Lie Groups,” Proc. of the 8th Control, Automation, Robotics and Vision Conference, 2004, Vol. 3, pp. 2211–2216, 2004.

[13] 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.

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

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

[17] G. H. Golub and C. F. Van Loan, “Matrix Computations,” The Johns Hopkins University Press, London, 1989.