Available online at www.sciencedirect.com
ScienceDirect
IFAC-PapersOnLine 49-22 (2016) 210–215

Distributed Visual 3-D Localization of A Human Using Pedestrian Detection Algorithm: A Passivity-Based Approach

Takeshi Hatanaka * **Riku Funada** * **Gülsüm Gezer** ** **Masayuki Fujita** *

* *Dept. Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology, 2-12-1 S5-16 O-okayama Meguro-ku, Tokyo 152-8550, Japan (e-mail: hatanaka@ctrl.titech.ac.jp)*
** *Dept. Control and Automation Eng., Yildiz Tech. Univ., Turkey*

Abstract: In this paper, we investigate a distributed visual 3-D localization of a human using a pedestrian detection algorithm for a camera network. We first formulate the problem as a distributed optimization problem. Then, PI consensus estimator-based distributed optimization scheme is extended so that it is applicable to the present problem. We then prove convergence to the optimal solution based on the passivity paradigm. We finally demonstrate the present solution through simulation both in static and dynamic cases.

Keywords: 3-D Human Localization, Pedestrian Detection Algorithm, Camera Network, Distributed Optimization, Passivity

1. INTRODUCTION

Owing to recent advances of pedestrian detection techniques in computer vision, well summarized in Dollar et al. (2012), the human activities are efficiently extracted from image data in real time without any expert knowledge. Indeed, such algorithms are already installed in MATLAB, OpenCV and other related software. In this paper, we address distributed 3-D localization of a human for a camera network using the pedestrian detection algorithms.

We first formulate the localization problem as a distributed optimization problem while modeling the human by an ellipsoid. We then present a distributed solution to the problem. Despite close relations to the distributed estimation as studied in Kamal et al. (2016); Olfati-Saber. (2007); Carli et al. (2007); Khan et al. (2010); Song et al. (2011); Hatanaka and Fujita (2013), the present problem formulation and solution are newly presented in this paper. The solution is also related to the consensus-based distributed optimization (Nedic and Ozdaglar (2009); Nedic et al. (2010); Zhu and Martinez (2012); Chang et al. (2014)). Nedic and Ozdaglar (2009) take a constant step size in the algorithm and quantify the error between the optimal solution and the time average of the estimates. However, the time-averaging is not always acceptable when the human to be localized is dynamic. While Nedic and Ozdaglar (2009) provides only an approximate solution, Nedic et al. (2010); Zhu and Martinez (2012); Chang et al. (2014) ensures exact convergence to the solution without using the time-averaging. However, they use a diminishing step size, which is also applicable only to a static problem.

On the other hand, Wang and Elia (2010, 2011) present a control theoretic solution to the above problem, which is expected to work even for the dynamic problem. However, exact convergence to the optimal solution is not achieved for nonlinear constraint functions. Since the constraints of our problem is convex but nonlinear, we thus present a new solution so that exact convergence is achieved. In the proof of convergence, we take an approach slightly different from Wang and Elia (2011), namely passivity-based approach. The above convergence result is obtained assuming well-definedness of the cost and constraints all over the vector space, but the cost of our problem is not always well-defined outside of the feasibility region, which may cause a problem since the estimates generated by the present solution, and Wang and Elia (2010, 2011) as well, are not ensured to remain in the region in the transient state. We thus prove that the estimates do never get into the ill-conditioned region.

We finally build a simulator on 3-D animation and graphics software in order to simulate not only the localization process but also image acquisition and processing. Then,the present solution is demonstrated for both static and dynamic cases on the simulator.

2. PROBLEM FORMULATION

Let us consider the situation where n cameras are distributed in the 3-D Euclidean space, and some of them capture a human, as illustrated in Fig. 1. In the sequel, the set of cameras which detect the human is denoted by $\mathcal{V}_h$. We denote the world frame and camera i-th frame by Σ_w and Σ_i, respectively. The position vector of the origin of Σ_i relative to Σ_w is denoted by $p_{wi} \in \mathbb{R}^3$, and the orientation of Σ_i relative to Σ_w is denoted by $R_{wi} \in SO(3)$. In this paper, we assume that the cameras are calibrated *a priori*, and p_{wi} and R_{wi} are available at each time instant.

In recent years, advanced pedestrian detection algorithms have been developed, which allows one to localize the

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.10.398

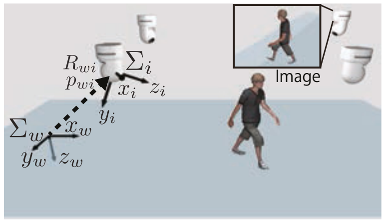

Fig. 1. Intended scenario: a human lives in the 3-D Euclidean space and is captured by a camera network.

human on the image in real time. A detection result is shown in Fig. 2, where the algorithm outputs the rectangle in which a human lives. We also denote the image plane coordinates of the points inside of the rectangle for camera i by $[x_{\min,i}, x_{\max,i}]\times[y_{\min,i}, y_{\max,i}]$, whose center is denoted by c_i. Then, the human must be inside of the following convex cone defined in the frame Σ_i.

$$\mathcal{H}_i := \left\{p \in \mathbb{R}^3 \mid M_i p \leq 0\right\},\ M_i := \begin{bmatrix} 1 & 0 & -\frac{x_{\max,i}}{\lambda_i} \\ -1 & 0 & \frac{x_{\min,i}}{\lambda_i} \\ 0 & 1 & -\frac{y_{\max,i}}{\lambda_i} \\ 0 & -1 & \frac{y_{\min,i}}{\lambda_i} \end{bmatrix}.$$

Transforming the coordinates from Σ_i to Σ_w(Hatanaka et al. (2015b)), the set is reformulated as $\mathcal{H}_{wi} := \left\{p_w \in \mathbb{R}^3 \mid M_{wi}p_w \leq m_{wi}\right\}$ with $M_{wi} := M_i R_{wi}^T$, $m_{wi} := M_i R_{wi}^T p_{wi}$.

In this paper, we model the human by an ellipsoid defined in Σ_w as $\Omega(Q, q) := \{p \in \mathbb{R}^3 |\ (p-q)^T Q^{-2}(p-q) \leq 1\}$, where $Q \in \mathbb{S}^{3\times3}$, and it contains six variables to be determined [1]. In the following, the map from $\mathbb{R}^6$ to $\mathbb{S}^{3\times3}$ is denoted by $\wedge$ and its inverse map is by $\vee$. Using the operator, we define the variables to be determined by $z := (q, Q^\vee) \in \mathbb{R}^9$. In the sequel, we denote the set of all 3×3 positive definite matrices by $\mathbb{S}_+^{3\times3}$ and define $\check{\mathbb{S}}_+^{3\times3} := \{x \in \mathbb{R}^6 |\ \hat{x} \in \mathbb{S}_+^{3\times3}\}$.

The set $\Omega(Q, q)$, say the human, must be included in the set $\mathcal{H}_{wi}$ whose condition is known to be formulated as

$$\|Q(M_{wi}^{(j)})^T\| + M_{wi}^{(j)} q - m_{wi}^{(j)} \leq 0,\ \forall j = 1, 2, 3, 4. \tag{1}$$

$$M_{wi}q - m_{wi} \leq 0, \tag{2}$$

where $M_{wi}^{(j)} \in \mathbb{R}^{1\times3}$ and $m_{wi}^{(j)} \in \mathbb{R}$ denote the j-th row of M_{wi} and j-th element of m_{wi}, respectively. It is easy to confirm that the left-hand sides of (1) and (2) are convex in z. The matrix inequality constraint $Q > 0$ is also known to be convex, and can be reduced to three scalar inequalities by applying the Schur complement twice [2].

Defining a function $g_i : \mathbb{R}^9 \to \mathbb{R}^{11}$ appropriately, the collection of (1), (2) and $Q > 0$ is compactly described as $g_i(z) \leq 0$. However, if camera i does not detect the human ($i \notin \mathcal{V}_h$), the constraints (1) and (2) cannot be obtained. In this case, we define $g_i : \mathbb{R}^9 \to \mathbb{R}^3$ by collecting only the constraints corresponding to $Q > 0$. This is why the output dimension of g_i is denoted by m_i in the sequel.

[1] $\mathbb{S}^{3\times3}$ denotes the set of all 3-by-3 symmetric matrices.

[2] Since the strict inequality constraint $Q > 0$ is inconvenient for the subsequent discussions, we instead impose a slightly stronger condition $Q > \epsilon I_3$ for a sufficiently small $\epsilon > 0$ in practice.

Fig. 2. Output of the pedestrian detection algorithm presented in Dalal and Triggs (2005).

We next define the cost function to be minimized. First, we try to maximize the volume of the ellipsoid in order that the actual human is inside of the ellipsoid. The requirement is described by the minimization of the function

$$f_i^{\text{volume}}(z) := -\log\det(Q) \tag{3}$$

which is known to be strictly convex in $Q^\vee \in \check{\mathbb{S}}_+^{3\times3}$. Regarding the center q, we assume that it should be close to the line connecting the focal center and the center of the rectangle. Since the position coordinates in Σ_w of the center of the rectangle is given as $c_{wi} := R_{wi}\left[c_i^T\ \lambda_i\right]^T + p_{wi}$(Hatanaka et al. (2015b)), this specification is reflected by the cost function

$$f_i^{\text{dist}}(z) := \left\|\left(I_3 - \frac{c_{wi}c_{wi}^T}{\|c_{wi}\|^2}\right)(p_{wi} - q)\right\|^2 \tag{4}$$

whose square root is equal to the distance between q and the line. Note that the function is convex in q but is not strictly convex. Also, the function f_i^{dist} depends on the measurement c_i, and hence cameras not included in $\mathcal{V}_h$ cannot formulate the function. Thus, we let such cameras take $f_i^{\text{dist}}(z) = 0$. Using the above functions, we define the private cost function of camera i by

$$f_i(z) := f_i^{\text{volume}}(z) + w_i f_i^{\text{dist}}(z), \tag{5}$$

where $w_i > 0$ is a weighting coefficient.

Let us formulate the optimization problem to be solved as

$$\min_{z\in\mathbb{R}^9} f(z) := \sum_{i=1}^n f_i(z) \text{ subject to } g_i(z) \leq 0\ \forall i. \tag{6}$$

It is easy to confirm that the functions f_i $(i = 1, 2, \ldots, n)$ are differentiable and convex if $Q \in \mathbb{S}_+^{3\times3}$, namely in the feasibility region of (6). Denoting the j-th element of g_i by $g_{ij}(j = 1, 2, \ldots, m_i)$, the functions g_{ij} $(j = 1, 2, \ldots, m_i,\ i = 1, 2, \ldots, n)$ are also convex. In addition, g_{ij} $(j = 1, 2, \ldots, m_i,\ i = 1, \ldots, n)$ are all differentiable at least in the feasibility region.

In this paper, we employ the following assumptions.

Assumption 1. $\cap_{i\in\mathcal{V}_h}\mathcal{H}_{wi}$ is compact and has an interior.

The former condition implicitly requires that the number of elements of the set $\mathcal{V}_h$ is greater than or equal to 2. Under the latter condition, the problem (6) satisfies the Slatter's constraint qualification, i.e., there exists z such that $g_i(z) < 0\ \ \forall i = 1, 2, \ldots, n$. In addition, the existence of the optimal solution to (6) and the finiteness of the minimal f are also guaranteed by the assumption.

Assumption 2. There exists a pair $(i, j) \in \mathcal{V}_h \times \mathcal{V}_h$ such that c_{wi} and c_{wj} are linearly independent of each other.