Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

WeC06.6

Constructive Camera Pose Control for Optimizing Multiview Distributed Video Coding

Jianghua Zhong[†‡], Xiaoming Hu[†], W. Bastiaan Kleijn[‡], and Ermin Kozica[‡]

Abstract— In this paper camera pose control for optimizing multiview distributed video coding is considered. The scenario considered is that multiple agents with monocular cameras observe a common scene in a three dimensional world. To get a good video reconstruction under a transmission rate constraint, the camera closest to the center of the camera array is chosen as the reference camera. The poses of all other cameras are controlled and designed such that their images are maximally similar under a constraint on their separation. Based on the rigid motion allowed for the cameras, two cases are considered. For the case where the rigid motion only involves translation, translation control is designed. For the case where it involves both translation and rotation, both controls are constructed. Some simulated results are given to show the efficiency of the designed controllers.

I. INTRODUCTION

Recently multiview imaging has attracted much attention due to the decreasing camera cost and its increasingly wide range of applications, such as in entertainment, security surveillance, and industry inspection. However, obtaining a good video reconstruction from multiview data is not always straightforward. First, the different internal parameters of cameras such as focal length, principal point and aspect ratio, and different external camera parameters such as position and orientation, lead to significantly different images, which makes it difficult to synthesize a virtual view. Second, the vast raw bit rate of multiview video requires efficient data compression, which may result in a high level of distortion. Fortunately, because the video data originate from the same scene, the inherent similarities of the multiview images can be exploited in their encoding. The similarities include in-terview similarities between images of adjacent cameras and temporal similarities between temporally successive images of each camera. An extensive literature exists in the area of multiview video compression, e.g., [9], [12], [3]. For an overview we refer to [4] and the references therein. The common idea of the existing results is to exploit the similarities (i.e., the correlations) between adjacent views in addition to the temporal and spatial similarities within a single view.

Distributed video coding is a relatively new paradigm for video compression that is based on information-theoretic results of Slepian and Wolf for the lossless case [18] and

Wyner and Ziv for the lossy case [24] -[27]. In distributed source coding, side information is considered to be available at the decoder but not at the encoder. Thus, the method exploits the source statistics in the decoder rather than in the encoder. As a result, The encoder has a low computational complexity, while the decoder has a high computational complexity. The input data is divided into cosets at the encoder and it transmits the index(i.e., the syndrome) of the coset. The receiver decodes by selecting the codeword in that coset that is most likely given the side information known to it. Therefore, the more similar the encoded data and the side information are, the better the reconstruction results under the constraint of transmission rate. Many results on distributed video coding exist see, e.g., [17], [13], the overview [5] and the references therein.

Due to the high data rate required for the transmission and storage at the encoder in multiview video, distributed video coding has been applied to lighten this burden [30], [23], [6], [10]. Multiview distributed video coding provides a technique for efficient multiview video compression. It requires only a low-complexity encoder at each camera.

In addition to the cited work on compression algorithms for video data, work on visual motion estimation is relevant to our problem considered in this paper. We can classify this work into point-based [20], line-based [22], [28], and curve-based or model based [1], [2], [21]. Furthermore, observers have been designed [14], [8] for the nonlinear dynamical model given in [20]. However, in contrast to our work, the model used in the latter papers is continuous time.

Lots of camera arrays has been also built for multiview imaging. For instance, Naemura et al. constructed a camera array system consisting of 16 cameras [16], while Zhang and Chen proposed a self-reconfigurable camera array with 48 cameras [29]. However, all existing results on designing camera array system only give system configurations, but not show how to control the camera by servo controllers.

This paper addresses the problem of control design for the camera pose(e.g., location and orientation) for the purpose of multiview distributed video coding. To the best knowledge of the authors, no earlier studies have been performed on this topic. Based on the nonlinear discrete-time dynamical model given in [20], we design a camera pose control that makes the images of all cameras maximally similar under a minimum distance constraint. The increased similarity of the images leads to better reconstruction results under the constraint of transmission rate.

The remainder of this paper is organized as follows. Section II presents the problem statement and Section III

gives some preliminaries. Section IV designs the camera pose controls. Based on the type of motion allowed for the camera, two cases are considered. One case is that the camera motion involves only translation, and the other is that its motion involves both translation and rotation. To validate the designed controllers, some simulations are given in Section V. Section VI provides the conclusion.
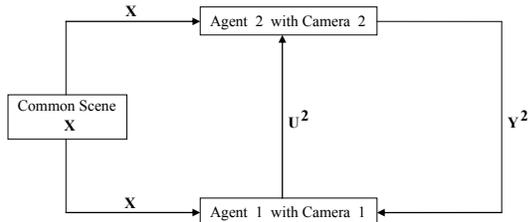
II. PROBLEM STATEMENT

Fig. 1. System description.

In this paper we consider the control of multiple agents with monocular cameras observing a common scene in a three-dimensional(3**D**) world to form a stereo view. We assume that all cameras have the same internal parameters. The best view reconstruction is possible when the observations of the cameras are maximally similar, as measured by an appropriate measure. However, due to the difference in the poses of the cameras, the images formed by them may be quite different. This leads to high distortion between the reconstruction and the original image under the constraint of transmission rate. To make their images maximally similar under the constraint that the distance between any two cameras is subject to a minimum distance constraint, we choose the camera that is closest to the center of the camera array, as the reference camera. The locations and/or orientations of all but the reference camera are corrected, by means of commands from the reference camera.

Without loss of generality, we can consider a communication system with two cameras, respectively attached to two agents, one of them being the reference camera. As depicted in Fig. 1, Camera 1 and Camera 2 are attached to Agent 1 and Agent 2, respectively. They observe a common scene **X** in a 3**D** world, and are connected by a wireless connection. Camera 1 is assumed to be the reference camera. We define

$$\mathbf{Y}^1 = \{Y_k^1\}_{k=1}^\infty, \ \mathbf{Y}^2 = \{Y_k^2\}_{k=1}^\infty, \ \mathbf{U}^2 = \{U_k^2\}_{k=1}^\infty, \quad (1)$$

where Y_k^1 and Y_k^2 are the images formed respectively by Camera 1 and Camera 2 at time instance k, and U_k^2 is a command to Agent 2 to correct the position and/or orientation of Camera 2, given by the Agent 1 at time instance k. For simplicity, we assume Agent 2 sends the image formed by Camera 2 to Agent 1, and the distortion between the image sent by Agent 2 and its corresponding reconstructed image by Agent 1 is negligible, and the rate for transmission of control information from Agent 1 to Agent 2 is negligible as well.

For the sake of simplicity, we restrict our attention to the case where the common scene is static and Camera 2 moves rigidly to change its location and/or orientation after receiving commands from Camera 1. Then the object of the static scene moves rigidly relative to Camera 2. Although the existing methods for the motion estimation, as mentioned in the previous section, may be classified as point-based, line-based, and curve-based or model based, Levoy and Whitted have pointed out that a discrete array of points arbitrarily displayed in space using a tabular array of perturbations can be rendered as a continuous three-dimensional surface, and proved that a wide class of geometrically defined objects, including both flat and curved surface, can be converted into points [11]. Therefore, here it is reasonable for us to study the case where the common scene is represented by N feature points. We assume the feature points in the images are available and their correspondences are known. To characterize the image of static scene in a 3**D** world, the camera model is crucial. Here we consider the most commonly used camera model, the pinhole(i.e., perspective projection) camera model.

In summary, the goal of this paper is to solve the following problem:

Problem P. *Use the available data* $\mathbf{Y}^1$ *and* $\mathbf{Y}^2$ *to design* $\mathbf{U}^2$ *such that* $\mathbf{Y}^2$ *is maximally similar to* $\mathbf{Y}^1$ *under the constraint that the corrected location of Camera 2 is at least a minimum length* d_{21} *away from Camera 1.*

III. PRELIMINARIES

In this section, we first revisit the pinhole camera model and recall some relevant results on visual motion estimation based on this camera model. Then we reduce Problem P stated in the previous section to a control design problem for a nonlinear control system.
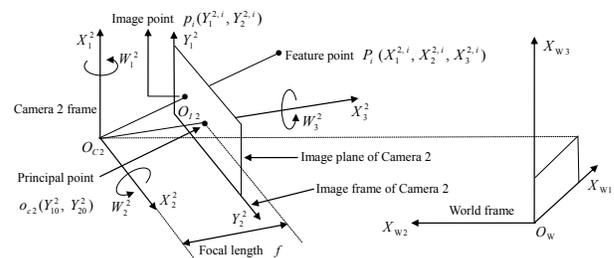

Fig. 2. Pinhole camera model of Camera 2.

The pinhole camera model is a nonlinear mapping from 3**D** world coordinates to 2**D** image coordinates. It gives an idealized mathematical framework, which is quite accurate for high quality camera systems [7]. Figure 2 depicts the pinhole model of Camera 2, where all coordinate frames are assumed to be right-handed orthogonal frames. The projection center is the origin O_{C2} of Camera 2 frame. The image plane of Camera 2, $Y_1^2 Y_2^2$, is parallel to the $X_1^2 X_2^2$ plane(the axes Y_1^2 and Y_2^2 are parallel to the axes X_1^2 and X_2^2 respectively), and it is displaced with the distance f(focal length) from O_{C2} along X_3^2-axis. The X_3^2-axis is also called

the optical axis, or the principal axis, and the intersection of the image plane of Camera 2 and the optical axis(here it is X_3^2-axis), is called the principal point(in the image plane of Camera 2) $o_{c2}(Y_{10}^2, \ Y_{20}^2)$.

For ease of notation, we first denote the set of natural numbers(positive integers) by $\mathbb{N}$, and the common Euclidean norm by $\| \cdot \|$ for a vector or a matrix, and $| \cdot |$ for a scalar. I indicates the identity matrix of three dimensions. For a square matrix A, $\exp(A)$ denotes the exponential function of A.

Let $V_k^2 = [V_{1,k}^2 \ V_{2,k}^2 \ V_{3,k}^2]^T$ indicate the translational velocity of the origin of Camera 2 frame at time instance k with respect to time instance $k+1$, and $W_k^2 = [W_{1,k}^2 \ W_{2,k}^2 \ W_{3,k}^2]^T$ denote the rotational velocity of Camera 2 frame at time instance k with respect to time instance $k+1$. Moreover, let

$$\Omega_k^2 = \begin{bmatrix} 0 & -W_{3,k}^2 & W_{2,k}^2 \\ W_{3,k}^2 & 0 & -W_{1,k}^2 \\ -W_{2,k}^2 & W_{1,k}^2 & 0 \end{bmatrix}. \qquad (2)$$

Then the rotation matrix R_k^2 of Camera 2 frame at time instance k with respect to time instance $k + 1$, and the translation vector T_k^2 of Camera 2 frame at time instance k with respect to time instance $k + 1$, can be defined, respectively, as follows:

1) if $\|\Omega_k^2\| \neq 0$, define

$$R_k^2 := \exp(\Omega_k^2), \ T_k^2 := \mathcal{T}^2(\Omega_k^2)V_k^2, \qquad (3)$$

where $\mathcal{T}^2(\Omega_k^2) := \frac{1}{\|\Omega_k^2\|}[(I - R_k^2)\Omega_k^2 + W_k^2(W_k^2)^T]$;

2) if $\|\Omega_k^2\| = 0$, define

$$R_k^2 := I, \ T_k^2 := V_k^2. \qquad (4)$$

We note that each rotation matrix R_k^2 is an orthogonal matrix with determinant 1.

Without loss of generality, we assume the focal length of Camera 2 is 1, its origin of image plane is its principal point, and its aspect ratio(i.e. the scaling in Y_2^2 direction divided by the scaling in the Y_1^2 direction) is 1. Then the positions of N feature points of the static scene with respect to Camera 2 frame and the projections of the feature points onto the image plane of Camera 2 can be expressed respectively as the following nonlinear dynamic system [19], [20]:

$$\begin{cases} X_{k+1}^{2,i} = R_k^2 X_k^{2,i} + T_k^2, \\ Y_k^{2,i} = [X_{1,k}^{2,i}/X_{3,k}^{2,i} \ \ X_{2,k}^{2,i}/X_{3,k}^{2,i}]^T + Z_k^{2,i}, \\ X_{3,k}^{2,i} \neq 0, \ i = 1, \ldots, N, \ \forall k \in \mathbb{N}, \end{cases} \qquad (5)$$

where the state $X_k^{2,i} = [X_{1,k}^{2,i} \ X_{2,k}^{2,i} \ X_{3,k}^{2,i}]^T \in \mathbb{R}^3 (X_{3,k}^{2,i}$ is called the depth of Point i at time instance k) denotes the position of Point i of the static scene, expressed in Camera 2 body-fixed frame at time instance k, the measurement output $Y_k^{2,i} = [Y_{1,k}^{2,i} \ Y_{2,k}^{2,i}]^T \in \mathbb{R}^2$(i.e., the images of Point i observed by Camera 2 at time instance k) indicates the position of Point i of the static scene, expressed in the image frame of Camera 2 at time instance k, and $Z_k^{2,i} = [Z_{1,k}^{2,i} \ Z_{2,k}^{2,i}]^T \in \mathbb{R}^2$ are measurement noises of Point i, measured by Camera 2 at time instance k, which is assumed

to be white, zero-mean, $(\delta_k^{2,i})^2$-variance, and Gaussian. The initial depth $X_{3,1}^{2,i}$ of each point i is assumed to be known.

Since all internal parameters of the cameras are assumed to be the identical, the positions of N feature points of the static scene with respect to Camera 1 frame and the projections of the feature points onto the image frame of Camera 1 can be expressed respectively as

$$\begin{cases} X_{k+1}^{1,i} = R_k^1 X_k^{1,i} + T_k^1, \\ Y_k^{1,i} = [X_{1,k}^{1,i}/X_{3,k}^{1,i} \ \ X_{2,k}^{1,i}/X_{3,k}^{1,i}]^T + Z_k^{1,i}, \\ X_{3,k}^{1,i} \neq 0, i = 1, \ldots, N, \ \forall k \in \mathbb{N}. \end{cases} \qquad (6)$$

As Camera 1 is the reference camera, we can assume:

Assumption A1 *The rigid motion of Camera 1 is known, that is, R_k^1 and T_k^1, $\forall k \in \mathbb{N}$, are known.*

We also make the following two assumptions:

Assumption A2 *The initial depth $X_{3,1}^{1,i}$ and initial image $Y_1^{1,i}$, $\forall i \in \{1, 2, \ldots, N\}$, are known.*

Assumption A3 *The measurement noises, $Z_k^{1,i}$, $i = 1, 2, \ldots, N$, $\forall k \in \mathbb{N}$, are white, zero-mean, $(\delta_k^{1,i})^2$-variance, and Gaussian, and independent of $[X_{1,k}^{1,i}/X_{3,k}^{1,i} \ \ X_{2,k}^{1,i}/X_{3,k}^{1,i}]^T$.*

Since the common scene is described by N feature points, their images formed by Camera 1 and Camera 2 at time instance k can be respectively expressed as $Y_k^1 = \{Y_k^{1,i}\}_{i=1}^N$, and $Y_k^2 = \{Y_k^{2,i}\}_{i=1}^N$, $\forall k \in \mathbb{N}$. Denote $Y^{1n} = \{Y_k^1\}_{k=1}^n$ and $Y^{2n} = \{Y_k^2\}_{k=1}^n$, $\forall n \in \mathbb{N}$. Then, to make $\mathbf{Y}^1$ and $\mathbf{Y}^2$ maximally similar under the distance constraint, is equivalent to choosing a proper conditional probability density function $f_{Y_k^{2,i}|Y_k^{1,i}}$ of $Y_k^{2,i}$ given $Y_k^{1,i}$ such that

$$\begin{aligned} \mathrm{E}[D(Y^{1n}, Y^{2n})] &= \sum_{k=1}^n \sum_{i=1}^N \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} [(y_{1,k}^{1,i} - y_{1,k}^{2,i})^2 \\ &+ (y_{2,k}^{1,i} - y_{2,k}^{2,i})^2] f_{Y_k^{2,i}|Y_k^{1,i}}(y_k^{2,i}|y_k^{1,i}) f_{Y_k^{1,i}}(y_k^{1,i}) \mathrm{d}y_k^{1,i} \mathrm{d}y_k^{2,i} \end{aligned}$$
$$(7)$$

is minimum. Here E is the expectation operator, $D(\cdot, \cdot)$ is the squared error distortion function, and $y_k^{j,i} = [y_{1,k}^{j,i} \ y_{2,k}^{j,i}]^T$, $j = 1, 2$, are realizations (i.e., particular values) of random variables $Y_k^{j,i}$.

Note that under Assumptions A1, A2, and A3, from system (6), the probability density function $f_{Y_{k+1}^{1,i}}$, $\forall i \in \{1, 2, \ldots, N\}$, $\forall k \in \mathbb{N}$, of the image $Y_{k+1}^{1,i}$ of Camera 1, is zero-mean, $(\delta_1^{1,i})^2 + (\delta_{k+1}^{1,i})^2$-variance, and Gaussian, while $f_{Y_1^{1,i}}$ is the Dirac delta function as $Y_1^{1,i}$ is known. Then it is easy to see that to minimize the expression of Equation (7), $f_{Y_{k+1}^{2,i}|Y_{k+1}^{1,i}}$, $k \in \mathbb{N}$, should be

$$f_{Y_{k+1}^{2,i}|Y_{k+1}^{1,i}}(y_{k+1}^{2,i}|y_{k+1}^{1,i}) = \delta(y_{k+1}^{1,i} - y_{k+1}^{2,i}), \qquad (8)$$

where $\delta(\cdot)$ denotes the Dirac delta function. Hence,

$$Y_{k+1}^{1,i} = Y_{k+1}^{2,i}, \ i = 1, \ldots, N, \ \forall k \in \mathbb{N}. \qquad (9)$$

Equation (9) shows that the position of each Point i of the static scene in the image frame of Camera 1 is the same as in the image frame of Camera 2. It is obvious that if Camera 2 is controlled to the same position and same orientation as those of Camera 1, then (9) holds naturally. In the following, we will design a proper controller such that

Camera 2 is controlled to a different location from Camera 1, and the distance between them(i.e., the distance between the origin of Camera 1 frame and the origin of Camera 2 frame) is no less than a given positive constant d_{21}. Note that at some time instance k, one translation vector T_k^2 and one rotational velocity W_k^2 control N feature points, and none of translation vectors and rotational velocities may satisfy (9) for the N points under the distance constraint. However, at each time instance $k+1$, we can always choose a translation vector and/or a rotational velocity such that the squared error distortion

$$
\begin{aligned}
&D(Y_{k+1}^1, Y_{k+1}^2) \\
&= \sum_{i=1}^{N} \left\{ (Y_{1,k+1}^{1,i} - Y_{1,k+1}^{2,i})^2 + (Y_{2,k+1}^{1,i} - Y_{2,k+1}^{2,i})^2 \right\}
\end{aligned} \quad (10)
$$

between Y_{k+1}^1 and Y_{k+1}^2 is minimized.

From [15], the position vector of the origin of Camera 1 frame from the origin of Camera 2 frame at time instance $k+1$ is

$$
q_{k+1}^{21} = X_{k+1}^{2,i} - R_{k+1}^{21} X_{k+1}^{1,i}, \ \forall i \in \{1,2,\ldots,N\}, \ \forall k \in \mathbb{N}, \quad (11)
$$

where R_{k+1}^{21} denotes the rotation matrix of Camera 1 frame relative to Camera 2 frame at time $k+1$. Controlling Camera 2 to the position where it is d_{21} units away from Camera 1 means that $\|q_{k+1}^{21}\| \geq d_{21}, \ \forall k \in \mathbb{N}$. To achieve this goal, we first denote by $R_k^{wj}, \ j = 1,2, \ \forall k \in \mathbb{N}$, the rotation matrices of Camera j frame relative to the world frame at time instance k. Then it is reasonable to assume:

Assumption A4 *The initial rotation matrices R_1^{w1} and R_1^{w2} are known.*

From [15], it follows that

$$
R_{k+1}^{21} = R_{k+1}^{2w} R_{k+1}^{w1} = (R_1^{w2})^{-1} \left[\prod_{l=1}^{k} R_l^2 \right] \left[\prod_{m=1}^{k} R_m^1 \right]^{-1} R_1^{w1}. \quad (12)
$$

From (5) and (6), we get

$$
\begin{aligned}
&X_{k+1}^{j,i} = (\prod_{v=1}^{k} R_{k+1-v}^j) X_1^{j,i} + \sum_{v=1}^{k-1} (\prod_{l=v+1}^{k} R_{k+1+v-l}^j) T_v^j + T_k^j, \\
&i = 1,2,\ldots,N, \ j = 1,2,
\end{aligned} \quad (13)
$$

where $X_1^{2,i} = [Y_{1,1}^{2,i} X_{3,1}^{2,i} \quad Y_{2,1}^{2,i} X_{3,1}^{2,i} \quad X_{3,1}^{2,i}]^T$, provided that $Z_k^{2,i} = 0, \ \forall k \in \mathbb{N}$.

Since the noise is not measurable, in this paper we will only foucs on the control design in Problem P for the nominal case where measurement noises of Camera 2 are set to zero. Although one can combine some existing filter such as an EKF with the control design that we present here, in general convergence is not guaranteed. It is our plan to study the stability issue in a separate paper.

Note that without loss of generality, in (11) we can choose Point $i = 1$. Then the goal of this paper is reduced to solving the following problem:

Problem P′. *Under Assumptions A1, A2, A3, and A4, using system (5) with known initial depth $X_{3,1}^{2,i}$ of each point i and $Z_k^{2,i} = 0 (i = 1,2,\ldots,N, \ \forall k \in \mathbb{N})$, and the available information $Y_k^{1,i}$ and $Y_1^{2,i}$ to design appropriate*

translation vector $T_k^2, \ \forall k \in \mathbb{N}$, and/or rotation matrix R_k^2 (or equivalently, $W_k^2 = [W_{1,k}^2, \ W_{2,k}^2, \ W_{1,k}^2]^T$), $\forall k \in \mathbb{N}$, in order to minimize the squared error distortion (10), under the minimum distance constraint

$$
\|q_{k+1}^{21}\| = \|X_{k+1}^{2,1} - R_{k+1}^{21} X_{k+1}^{1,1}\| \geq d_{21}, \ \forall k \in \mathbb{N}, \quad (14)
$$

where R_{k+1}^{21} is of form of (12) and

$$
\begin{aligned}
&X_{k+1}^{j,1} = (\prod_{v=1}^{k} R_{k+1-v}^j) X_1^{j,1} \\
&+ \sum_{v=1}^{k-1} (\prod_{l=v+1}^{k} R_{k+1+v-l}^j) T_v^j + T_k^j, \ j = 1,2
\end{aligned} \quad (15)
$$

with $X_1^{2,i} = [Y_{1,1}^{2,i} X_{3,1}^{2,i} \quad Y_{2,1}^{2,i} X_{3,1}^{2,i} \quad X_{3,1}^{2,i}]^T$.

IV. POSE CONTROL DESIGN

In this section we solve Problem P′. In terms of the rigid motion allowed for the camera, here we consider two cases. One is that the rigid motion of Camera 2 only involves translation, which is referred to as Case 1 for ease of statement. The other is that it involves both translation and rotation about X_1^2-axis on $X_2^2 X_3^2$ plane, as depicted in Fig. 2, which is referred to as Case 2, similarly. For Case 1, the rotational velocity is identically zero, then according to the definition of rotation matrix in previous section, $R_k^2 \equiv I$, $\forall k \in \mathbb{N}$, while for Case 2, the rotational velocity of Camera 2 at time instance k with respect to time instance $k+1$, is $W_k^2 = [W_{1,k}^2 \ 0 \ 0]^T$, then

$$
R_k^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos W_{1,k}^2 & -\sin W_{1,k}^2 \\ 0 & \sin W_{1,k}^2 & \cos W_{1,k}^2 \end{bmatrix}, \ \forall k \in \mathbb{N}. \quad (16)
$$

It is seen that Case 1 is a special one of Case 2. In the following, we only give a detailed construction of the controller for Case 1, while directly follow a formula of the controller for Case 1 from Case 2.

For Case 2, from (16), system (5) without noise becomes the following nonlinear control system:

$$
\begin{cases}
X_{1,k+1}^{2,i} = X_{1,k}^{2,i} + T_{1,k}^2, \\
X_{2,k+1}^{2,i} = X_{2,k}^{2,i} \cos W_{1,k}^2 - X_{3,k}^{2,i} \sin W_{1,k}^2 + T_{2,k}^2, \\
X_{3,k+1}^{2,i} = X_{2,k}^{2,i} \sin W_{1,k}^2 + X_{3,k}^{2,i} \cos W_{1,k}^2 + T_{3,k}^2, \\
Y_k^{2,i} = [X_{1,k}^{2,i}/X_{3,k}^{2,i} \quad X_{2,k}^{2,i}/X_{3,k}^{2,i}]^T, \\
X_{3,k}^{2,i} \neq 0, \ i = 1,\ldots,N, \ \forall k \in \mathbb{N}
\end{cases} \quad (17)
$$

where the control input is $(T_{1,k}^2, T_{2,k}^2, T_{3,k}^2, W_{1,k}^2)$, and the initial depth $X_{3,1}^{2,i}$ of each point i is known.

For system (17), the squared error distortion (10) becomes

$$
\begin{aligned}
&d(Y_{k+1}^1, Y_{k+1}^2) \\
&= [T_{1,k}^2]^2 \sum_{i=1}^{N} \frac{1}{(X_{3,k+1}^{2,i})^2} + 2T_{1,k}^2 \sum_{i=1}^{N} \frac{1}{(X_{3,k+1}^{2,i})^2} \\
&\cdot [X_{1,k}^{2,i} - X_{3,k+1}^{2,i} Y_{1,k+1}^{1,i}] + [T_{2,k}^2]^2 \sum_{i=1}^{N} \frac{1}{(X_{3,k+1}^{2,i})^2} \\
&+ 2T_{2,k}^2 \sum_{i=1}^{N} \frac{1}{(X_{3,k+1}^{2,i})^2} [X_{2,k}^{2,i} \cos W_{1,k}^2 - X_{3,k}^{2,i} \sin W_{1,k}^2 \\
&- X_{3,k+1}^{2,i} Y_{2,k+1}^{1,i}] + \sum_{i=1}^{N} \frac{1}{(X_{3,k+1}^{2,i})^2} \left\{ [X_{1,k}^{2,i} - X_{3,k+1}^{2,i} Y_{1,k+1}^{1,i}]^2 \right. \\
&\left. + [X_{2,k}^{2,i} \cos W_{1,k}^2 - X_{3,k}^{2,i} \sin W_{1,k}^2 - X_{3,k+1}^{2,i} Y_{2,k+1}^{1,i}]^2 \right\}.
\end{aligned} \quad (18)
$$

Note that $T_{3,k}^2$ is relative to $X_{3,k+1}^{2,i}$ with the form of the third equation in (17). So the distortion function in (18) is nonlinear with respect to the variable $T_{3,k}^2$. Obviously, it is also nonlinear with respect to the variable $W_{1,k}^2$. Hence, the problem of minimizing (18) becomes quite difficult under the distance constraint (14) and constraint $X_{3,k+1}^{2,i} \neq 0$, $\forall i \in \{1, 2, \ldots, N\}$, and in general there is no analytic optimal solution for this constrained nonlinear optimization problem. However, it is easy to see that if the optimal $T_{3,k}^2$ and $W_{1,k}^2$, $\forall k \in \mathbb{N}$, are obtained(i.e., the optimal $X_{3,k+1}^{2,i}$, $\forall i \in \{1, 2, \ldots, N\}$ is known), then from (18), it is not difficult to get the optimal $T_{1,k}^2$ and $T_{2,k}^2$, $\forall k \in \mathbb{N}$, under the distance constraint (14) and constraint $X_{3,k+1}^{2,i} \neq 0 (\forall i \in \{1, 2, \ldots, N\}, \forall k \in \mathbb{N})$. Hence, to deal with this constrained nonlinear optimization problem, we solve it in three steps as follows.

Step 1. Get the optimal $T_{3,k}^2$, $\forall k \in \mathbb{N}$, under distance constraint (14) and the constraint $X_{3,k+1}^{2,i} \neq 0$, $\forall i \in \{1, 2, \ldots, N\}, \forall k \in \mathbb{N}$, when Camera 2 only translates along X_3^2-axis, using a linearization approach.

Step 2. Get the optimal $W_{1,k}^2$, $\forall k \in \mathbb{N}$, under the constraint $X_{3,k+1}^{2,i} \neq 0$, $\forall i \in \{1, 2, \ldots, N\}, \forall k \in \mathbb{N}$, based on the obtained optimal $T_{3,k}^2$ in Step 1, when Camera 2 only rotates about X_1^2-axis, using the linearization approach as well.

Step 3. Get the optimal $T_{1,k}^2$, and $_{2,k}^2$, $\forall k \in \mathbb{N}$, under the distance constraint (14), based on the obtained optimal $T_{3,k}^2$ in Step 1 and the optimal $W_{1,k}^2$ in Step 2.

Here the linearization approach means: for an objective function

$$\sum_{i=1}^{N} \left\{ [a_i - f_i(x, y)]^2 + [b_i - g_i(x, y)]^2 \right\}, \quad (19)$$

where the analytic functions $f_i(x, y)$ and $g_i(x, y)$ are nonlinear with respect to the variable x, we replace the nonlinear functions $f_i(x, y)$ and $g_i(x, y)$ respectively by their first approximate linearizations

$$f_i(0, y) + x \frac{\partial f_i}{\partial x}|_{x=0} := f_i(0, y) + f_x'(0, y)x,$$
$$g_i(0, y) + x \frac{\partial g_i}{\partial x}|_{x=0} := g_i(0, y) + g_x'(0, y)x,$$

then the objective function (19) are approximately equal to

$$\sum_{i=1}^{N} \left\{ [a_i - f(0, y) - f_x'(0, y)x]^2 + [b_i - g_i(0, y) - g_x'(0, y)x]^2 \right\}. \quad (20)$$

Clearly, getting the optimal x_{opt} is not difficult from the objective function (20) under the distance constraint (14) and the constraint $X_{3,k+1}^{2,i} \neq 0 (\forall i \in \{1, 2, \ldots, N\}, \forall k \in \mathbb{N}$. But this x_{opt} may be a suboptimal solution for optimizing the objective function (19). The error between the x_{opt} and the actual optimal solution depends on the solution approach. In the next section, we will strive to get a better suboptimal solution and make the error smaller.

In the following, we will deal with the three steps one by one. First, we strive to solve Step 1.

When Camera 2 only translates along X_3^2-axis, but not rotate about any axes, the system (5) without measurement

noise becomes

$$\begin{cases} X_{1,k+1}^{2,i} = X_{1,1}^{2,i}, \\ X_{2,k+1}^{2,i} = X_{2,1}^{2,i}, \\ X_{3,k+1}^{2,i} = X_{3,k}^{2,i} + T_{3,k}^2, \\ Y_k^{2,i} = [X_{1,k}^{2,i}/X_{3,k}^{2,i} \quad X_{2,k}^{2,i}/X_{3,k}^{2,i}]^T, \\ X_{3,k}^{2,i} \neq 0, \ i = 1, \ldots, N, \ \forall k \in \mathbb{N}, \end{cases} \quad (21)$$

For ease of statement, we denote

$$\bar{X}_{k+1}^{21} = [\bar{X}_{1,k+1}^{21} \ \bar{X}_{2,k+1}^{21} \ \bar{X}_{3,k+1}^{21}]^T := R_{k+1}^{21} X_{k+1}^{1,1} \quad (22)$$

with R_{k+1}^{21} in the form of (12) and $X_{k+1}^{1,1}$ in the form of (13) as $i = 1$ and $j = 1$. Note that in this case $R_k^2 \equiv I$, $\forall k \in \mathbb{N}$. Then from (12), it yields $R_{k+1}^{21} = [R_1^{w2}]^{-1}[\prod_{m=1}^{k} R_m^1]^{-1} R_1^{w1}$. Hence according to the assumptions, the probability density function of $\bar{X}_{k+1}^{21}$, which is of the form of (22), is known. Then the distance constraint (14) becomes

$$(d_{21})^2 \leq [X_{1,1}^{2,1} - \bar{X}_{1,k+1}^{21}]^2 + [X_{2,1}^{2,1} - \bar{X}_{2,k+1}^{21}]^2 + [X_{3,k}^{2,1} + T_{3,k}^2 - \bar{X}_{3,k+1}^{21}]^2. \quad (23)$$

Define

$$d_{k+1} := (d_{21})^2 - [X_{1,1}^{2,1} - \bar{X}_{1,k+1}^{21}]^2 - [X_{2,1}^{2,1} - \bar{X}_{2,k+1}^{21}]^2 \quad (24)$$

with $X_{1,1}^{2,1} = X_{3,1}^{2,1} Y_{1,1}^{2,1}$ and $X_{2,1}^{2,1} = X_{3,1}^{2,1} Y_{2,1}^{2,1}$. Then solving Inequality (23) yields:

1) if $d_{k+1} \leq 0$, then any $T_{3,k}^2 \in \mathbb{R}$ satisfies (23);
2) if $d_{k+1} > 0$, then

$$T_{3,k}^2 \geq \sqrt{d_{k+1}} + \bar{X}_{3,k+1}^{21} - X_{3,k}^{2,1} := L_k. \quad (25)$$

Denote

$$a_{1,k}^{2,i} = \frac{X_{1,k+1}^{2,i}}{X_{3,k+1}^{2,i}}|_{T_{3,k}^2=0}, \quad a_{2,k}^{2,i} = \frac{X_{2,k+1}^{2,i}}{X_{3,k+1}^{2,i}}|_{T_{3,k}^2=0},$$
$$b_{1,k}^{2,i} = \frac{\partial \left[\frac{X_{1,k+1}^{2,i}}{X_{3,k+1}^{2,i}}\right]}{\partial T_{3,k}^2}|_{T_{3,k}^2=0}, \quad b_{2,k}^{2,i} = \frac{\partial \left[\frac{X_{2,k+1}^{2,i}}{X_{3,k+1}^{2,i}}\right]}{\partial T_{3,k}^2}|_{T_{3,k}^2=0}, \quad (26)$$

where $X_{1,k+1}^{2,i}$, $X_{2,k+1}^{2,i}$ and $X_{3,k+1}^{2,i}$ are of the forms of (21).

It is noticeable that $a_{1,k}^{2,i}$, $a_{2,k}^{2,i}$, $b_{1,k}^{2,i}$ and $b_{2,k}^{2,i}$ defined in (26) are only relative to $X_{1,1}^{2,i}$, $X_{2,1}^{2,i}$, $X_{3,1}^{2,i}$, and $T_{3,1}^2, \ldots, T_{3,k-1}^2$, which are the variables before time instance k.

Define

$$\lambda_{3,k}^2 = \frac{1}{\sum_{i=1}^{N} \left\{[b_{1,k}^{2,i}]^2 + [b_{2,k}^{2,i}]^2\right\}} \sum_{i=1}^{N} \left\{ (b_{1,k}^{2,i} [Y_{1,k+1}^{1,i} - a_{1,k}^{2,i}] + b_{2,k}^{2,i}[Y_{2,k+1}^{1,i} - a_{2,k}^{2,i}] \right\}, \ \forall k \in \mathbb{N}. \quad (27)$$

By the linearization approach described before, we can get the optimal $T_{3,k}^2$, $\forall k \in \mathbb{N}$, under the constraint $X_{3,k+1}^{2,i} \neq 0$, $\forall i \in \{1, 2, \ldots, N\}$, using the following algorithm.

Algorithm 1. *For any $k \in \mathbb{N}$,*
1) *if $d_{k+1} \leq 0$, then*
   a) *if $|\lambda_{3,k}^2| \leq 1$, then $T_{3,k}^2 = \lambda_{3,k}^2$,*
   b) *if $\lambda_{3,k}^2 > 1$, then $T_{3,k}^2 = 1$,*
   c) *if $\lambda_{3,k}^2 < -1$, then $T_{3,k}^2 = -1$;*

2) if $d_{k+1} > 0$, then

    a) if $L_k \geq -1$, then $T_{3,k}^2 = \max\{L_k, \lambda_{3,k}^2\}$,

    b) if $L_k < -1$, then

        i) if $|\lambda_{3,k}^2| \leq 1$, then $T_{3,k}^2 = \lambda_{3,k}^2$,

        ii) if $\lambda_{3,k}^2 > 1$, then $T_{3,k}^2 = 1$,

        iii) if $\lambda_{3,k}^2 < -1$, then $T_{3,k}^2 = -1$;

if $X_{3,1}^{2,i} + \sum_{j=1}^{k-1} T_{3,j}^2 + T_{3,k}^2 = 0$ for some $i \in \{1, 2, \ldots, N\}$, there must exist some constant $0 < c_k < 1$ such that $X_{3,1}^{2,i} + \sum_{j=1}^{k-1} T_{3,j}^2 + \bar{T}_{3,k}^2 \neq 0$ for any $i \in \{1, 2, \ldots, N\}$, with

$$\bar{T}_{3,k}^2 = T_{3,k}^2 + \min\{|1 - \lambda_{3,k}^2|, |1 + \lambda_{3,k}^2|, c_k\}, \quad (28)$$

then we take $T_{3,k}^2 = \bar{T}_{3,k}^2$ as the optimal solution instead.

Now we deal with Step 2. Since we try to get the optimal $W_{1,k}^2$, based on the optimal $T_{3,k}^2$ (the optimal $T_{1,k}^2$ and $T_{2,k}^2$ have not obtained yet) in Step 2, we use the control system with $T_{1,k}^2 = T_{2,k}^2 \equiv 0$, $\forall k \in \mathbb{N}$, and the obtained optimal $T_{3,k}^2$ in Step 1. Thus we consider the following system:

$$\begin{cases} X_{1,k+1}^{2,i} = X_{1,k}^{2,i}, \\ X_{2,k+1}^{2,i} = X_{2,k}^{2,i} \cos W_{1,k}^2 - X_{3,k}^{2,i} \sin W_{1,k}^2, \\ X_{3,k+1}^{2,i} = X_{2,k}^{2,i} \sin W_{1,k}^2 + X_{3,k}^{2,i} \cos W_{1,k}^2 + T_{3,k}^2, \\ Y_k^{2,i} = [X_{1,k}^{2,i}/X_{3,k}^{2,i} \quad X_{2,k}^{2,i}/X_{3,k}^{2,i}]^T, \\ X_{3,k}^{2,i} \neq 0, \ i = 1, \ldots, N, \ \forall k \in \mathbb{N}, \end{cases} \quad (29)$$

where $T_{3,k}^2$ is the optimal translation along X_3^2-axis obtained in Step 1, and $W_{1,k}^2$ is the control input.

Denote

$$h_{1,k}^{2,i} = \frac{X_{1,k+1}^{2,i}}{X_{3,k+1}^{2,i}}|_{W_{1,k}^2 = 0}, \ h_{2,k}^{2,i} = \frac{X_{2,k+1}^{2,i}}{X_{3,k+1}^{2,i}}|_{W_{1,k}^2 = 0},$$

$$g_{1,k}^{2,i} = \frac{\partial \left[\frac{X_{1,k+1}^{2,i}}{X_{3,k+1}^{2,i}}\right]}{\partial W_{1,k}^2}|_{W_{1,k}^2 = 0}, \ g_{2,k}^{2,i} = \frac{\partial \left[\frac{X_{2,k+1}^{2,i}}{X_{3,k+1}^{2,i}}\right]}{\partial W_{1,k}^2}|_{W_{1,k}^2 = 0}, \quad (30)$$

where $X_{1,k+1}^{2,i}$, $X_{2,k+1}^{2,i}$ and $X_{3,k+1}^{2,i}$ are in forms of (29).

We also define

$$\omega_{1,k}^2 = \frac{1}{\sum_{i=1}^N \left\{[g_{1,k}^{2,i}]^2 + [g_{2,k}^{2,i}]^2\right\}} \sum_{i=1}^N \left\{ g_{1,k}^{2,i}[Y_{1,k+1}^{1,i} - h_{1,k}^{2,i}] + g_{2,k}^{2,i}[Y_{2,k+1}^{1,i} - h_{2,k}^{2,i}] \right\}, \ k \in \mathbb{N}. \quad (31)$$

Then similar to Case 1, we can get the optimal $W_{1,k}^2$ by the algorithm below.

Algorithm 2. For any $k \in \mathbb{N}$, if $W_{1,k}^2 = \omega_{1,k}^2$ and the designed $T_{3,k}^2$ in Algorithm 1 make $X_{3,k+1}^{2,i} = 0$ for some $i \in \{1, 2 \ldots, N\}$, with $X_{3,k+1}^{2,i}$ in the form of the third equation of system (29), then there must exist some constant $0 < w_k < 0.01$ such that $X_{3,k+1}^{2,i} \neq 0$ for any $i \in \{1, 2 \ldots, N\}$ when $W_{1,k}^2 = \alpha_{1,k}^2$ instead, with $\alpha_{1,k}^2 = \omega_{1,k}^2 + w_k$. In this case we choose $W_{1,k}^2 = \alpha_{1,k}^2$ as the optimal rotational velocity. Otherwise, we select $W_{1,k}^2 = \omega_{1,k}^2$ as the optimal one.

Finally, we deal with Step 3. From the squared error distortion function (18), it is easy to see that getting the optimal $T_{1,k}^2$ and $T_{2,k}^2$ is a optimization problem of minimizing a quadratic function outsides a circle. To solve it, we first give the following Lemma, which is straightforward.

Lemma 1. Let function $F(x, y) = ax^2 + bx + ay^2 + cy$ with $a > 0$. Then under the constraint $(x - x_0)^2 + (y - y_0)^2 \geq (d_0)^2$ with $d_0 > 0$, the minimum point (x_{min}, y_{min}) that minimizes the function F is expressed as follows.

1) If $(x_0 + \frac{b}{2a})^2 + (y_0 + \frac{c}{2a})^2 \geq (d_0)^2$, then $(x_{min}, y_{min}) = (-\frac{b}{2a}, -\frac{c}{2a})$.

2) If $(x_0 + \frac{b}{2a})^2 + (y_0 + \frac{c}{2a})^2 < (d_0)^2$, then

    a) if $x_0 = -\frac{b}{2a}$, and $y_0 = -\frac{c}{2a}$, then any $(x, y) \in \mathcal{S} := \{(x, y) | (x - x_0)^2 + (y - y_0)^2 = (d_0)^2\}$ can be chosen as (x_{min}, y_{min});

    b) if $x_0 = -\frac{b}{2a}$, and $y_0 \neq -\frac{c}{2a}$, then $x_{min} = x_0$, and

        i) if $y_0 > -\frac{c}{2a}$, then $y_{min} = y_0 - d_0$,

        ii) if $y_0 < -\frac{c}{2a}$, then $y_{min} = y_0 + d_0$;

    c) if $x_0 \neq -\frac{b}{2a}$, and $y_0 = -\frac{c}{2a}$, then $y_{min} = y_0$, and

        i) if $x_0 > -\frac{b}{2a}$, then $x_{min} = x_0 - d_0$,

        ii) if $x_0 < -\frac{b}{2a}$, then $x_{min} = x_0 + d_0$;

    d) if $x_0 \neq -\frac{b}{2a}$, and $y_0 \neq -\frac{c}{2a}$, then

        i) if $x_0 > -\frac{b}{2a}$, then $x_{min} = x_0 - d_0/\sqrt{1 + s^2}$ with $s = (2ay_0 + c)/(2ax_0 + b)$,

        ii) if $x_0 < -\frac{b}{2a}$, then $x_{min} = x_0 + d_0/\sqrt{1 + s^2}$,

        iii) if $y_0 > -\frac{c}{2a}$, then $y_{min} = y_0 - |s|d_0/\sqrt{1 + s^2}$,

        iv) if $y_0 < -\frac{c}{2a}$, then $y_{min} = y_0 + |s|d_0/\sqrt{1 + s^2}$.

In the function F, we set

$$\begin{cases} x = T_{1,k}^2, \ y = T_{2,k}^2, \ a = \sum_{i=1}^N \frac{1}{(X_{3,k+1}^{2,i})^2}, \\ b = 2\sum_{i=1}^N \frac{X_{1,k}^{2,i} - X_{3,k+1}^{2,i} Y_{1,k+1}^{1,i}}{(X_{3,k+1}^{2,i})^2}, \\ c = 2\sum_{i=1}^N \frac{X_{2,k}^{2,i}\cos W_{1,k}^2 - X_{3,k}^{2,i}\sin W_{1,k}^2 - X_{3,k+1}^{2,i} Y_{2,k+1}^{1,i}}{(X_{3,k+1}^{2,i})^2}, \\ x_0 = X_{1,k}^{2,1} - \bar{X}_{1,k+1}^{21}, \\ y_0 = X_{2,k}^{2,1}\cos W_{1,k}^2 - X_{3,k}^{2,i}\sin W_{1,k}^2 - \bar{X}_{2,k+1}^{21}, \\ d_0 = (d_{21})^2 - (X_{3,k+1}^{2,1} - \bar{X}_{3,k+1}^{21})^2. \end{cases} \quad (32)$$

Note that in this case the variables $X_{1,k}^{2,i}$, $X_{2,k}^{2,i}$, $X_{3,k+1}^{2,i}$ in (32) are of forms of (17) with the obtained optimal $T_{3,k}^2$ in Step 1 and the optimal $W_{1,k}^2$ in Step 2.

Hence, we get the following result.

Proposition 1. *The controller designed Algorithms 1, 2, and Lemma 1 with (32) and (17) solves Problem P′ when Camera 2 can rotate about X_1^2-axis and translate as well.*

Based on the results for Case 2, in the following, we directly follow the results for Case 1.

As mentioned before, in Case 1, $W_{1,k}^2 \equiv 0$ and $R_k^2 \equiv 0$, $\forall k \in \mathbb{N}$. Then in Case 1, system (5) without noise becomes

the following nonlinear control system:

$$\begin{cases} X_{1,k}^{2,i} = X_{1,k}^{2,i} + T_{1,k}^2, \\ X_{2,k}^{2,i} = X_{2,k}^{2,i} + T_{2,k}^2, \\ X_{3,k}^{2,i} = X_{3,k}^{2,i} + T_{3,k}^2, \\ Y_k^{2,i} = [X_{1,k}^{2,i}/X_{3,k}^{2,i},\ X_{2,k}^{2,i}/X_{3,k}^{2,i}]^T, \\ X_{3,k}^{2,i} \neq 0,\ i = 1,2,\ldots,N,\ k \in \mathbb{N}, \end{cases} \quad (33)$$

where $(T_{1,k}^2,\ T_{2,k}^2,\ T_{3,k}^2)$ is the control input, and the initial depth $X_{3,1}^{2,i}$ of each point i is assumed to be known.

In (32), we set $W_{1,k}^2 \equiv 0,\ \forall k \in \mathbb{N}$. Then together with Lemma 1 and (33), we can get the optimal $T_{1,k}^2, T_{2,k}^2, \forall k \in \mathbb{N}$ for Case 1. Therefore, we have the following result.

Proposition 2. *The controller designed in Algorithm 1, and Lemma 1 with (33) and (32), where $W_{1,k}^2 \equiv 0,\ \forall k \in \mathbb{N}$, solves Problem P' when Camera 2 can only translate.*

Remark 1. The above designed controllers for the two cases are recursive expressions at time instance $k(\forall k \in \mathbb{N})$, based on the depths and images of two cameras at initial time instance $k = 1$, and the images of Camera 1 from time instance $k = 1$ to $k + 1$. If the images formed by Camera 2 are required to transmit to Camera 1 as a feedback at time instance k_0, we can reset the initial time instance is $k = k_0$, then using the above designed controller we get a control formula at any time instance $k \geq k_0$.

Remark 2. For M cameras, we can control them one by one to assure that the distance between any two cameras is no less than a certain selected distance. Without loss of generality, we assume Camera 1 is the reference camera. First we control Camera 2 to a location where is d_{21} units away from Camera 1. For Camera 3, we select appropriate minimum distances d_{31} and d_{32}, and replace Inequality (14) by $\|q_{k+1}^{31}\| \geq d_{31}$ and $\|q_{k+1}^{32}\| \geq d_{32}$. For Camera j, $4 \leq j \leq M$, we choose proper minimum distances d_{jj_1}, $1 \leq j_1 \leq j - 1$, and replace Inequality (14) by $\|q_{k+1}^{jj_1}\| \geq d_{jj_1}$. Thus, all M cameras are controlled to different locations and the distance between any two cameras is subject to a minimum distance constraint. The most straightforward case is where all the constraint minimum distances are identical.

V. Simulated Results

In this section, due to limited space, we only give some simulations for Case 2 to show the efficiency of our designed controllers.

As in the previous sections, we assume Camera 1 is the reference camera, while Camera 2 is a controlled one. All measurement noises of two cameras are assumed to be negligible. We consider the case when Camera 1 remains static with respect to the world frame.

Using the notations introduced in Section III, we use as minimum distance constraint $d_{21} = 14$ and the initial time instance is $k = 1$. Moreover, we select the initial rotation matrices of Camera 1 frame and Camera 2 frame with respect to the world frame respectively are

$$R_1^{w1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}, \quad R_1^{w2} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} & 0 \\ -\frac{4}{5} & \frac{3}{5} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (34)$$

We consider the case where the static scene is represented by three feature points P_1, P_2, P_3. Their images are denoted by $\odot$, $\oplus$, $\otimes$ for Camera 1, and $\bullet$, $+$, $\times$ for Camera 2, respectively. Moreover, their initial depths in the frames of Cameras 1 and 2, respectively are $X_{3,1}^{1,1} = 1$, $X_{3,1}^{1,2} = -4$, $X_{3,1}^{1,3} = 5\sqrt{3}+1$, and $X_{3,1}^{2,1} = 1$, $X_{3,1}^{2,2} = 1$, $X_{3,1}^{2,3} = 11$. Then straightforward computations show that $\|q_1^{21}\| = 30$. Thus it is seen that the positions and orientations of two cameras differ significantly.

The initial images of the three points P_1, P_2, P_3 are

$$Y_1^{1,1} = \begin{bmatrix} -20 \\ 0 \end{bmatrix},\ Y_1^{1,2} = \begin{bmatrix} \frac{15}{2} \\ -\frac{5\sqrt{3}}{4} \end{bmatrix},\ Y_1^{1,3} = \begin{bmatrix} \frac{-30}{5\sqrt{3}+1} \\ \frac{5}{5\sqrt{3}+1} \end{bmatrix},$$

for Camera 1, and

$$Y_1^{2,1} = [6 \quad 8]^T,\ Y_1^{2,2} = [-8 \quad 6]^T,\ Y_1^{2,3} = [0 \quad 0]^T,$$

for Camera 2. They are shown in Fig. 3. It is seen that they are different due to the different positions and orientations of the two cameras, as stated above.

Using the previously designed controllers, we get the squared error distortions and controlled images for Case 2, which are shown respectively in Fig. 4 and Fig. 5.

Fig. 4 shows that the distortion between the initial images is high, but it decreases significantly after Camera 2 is controlled and always decreases as Camera 2 is controlled persistently, while it decreases slightly after the third time instance. However, due to the distance constraint between two cameras, there is still lower distortion between two cameras' images even if Camera 2 has been controlled nine times.
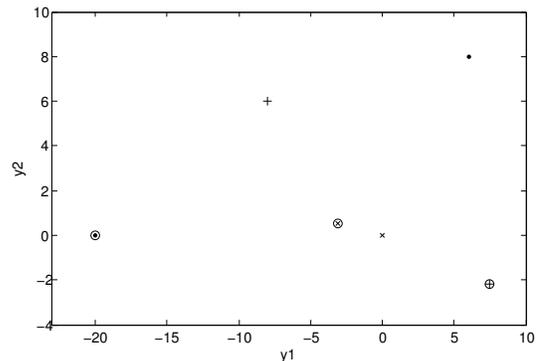

Fig. 3.   Images at initial time instance $k = 1$.

VI. Conclusion

We considered control for multiview distributed video coding with a rate constraint. The reconstruction accuracy improves if the images of adjacent cameras are more similar. So it is important for multiview distributed video coding to have the images of all cameras maximally similar under the constraint that any camera is at least at a minimum distance away from the others. In this paper we developed the control design for camera pose to make the images of all cameras maximally similar under the distance constraint. Using rigid
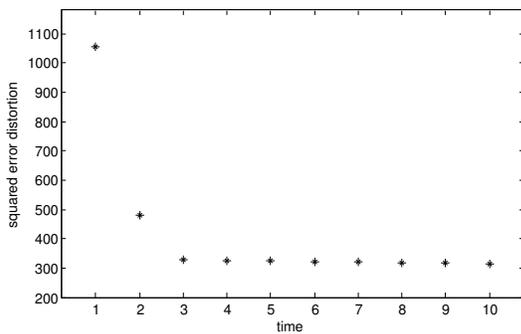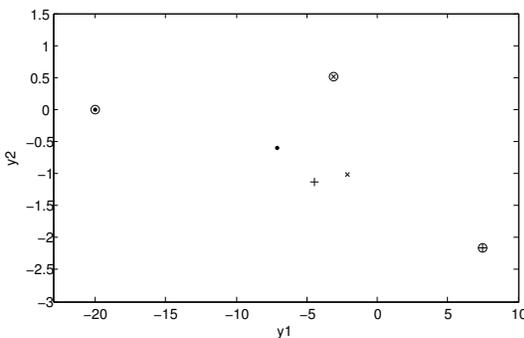
Fig. 4.   Squared error distortion for Case 2.

Fig. 5.   Images at time instance $k = 10$ for Case 2.

motion of the camera, two cases were considered. One is that the rigid motion of the camera only involves translation. The other is that its rigid motion involves both translation and rotation. For both cases, we gave an appropriate controller that minimizes the squared error distortion. Some simulated results were provided to validate our designed controllers.

REFERENCES

[1] Arbogast, E., and Mohr, R.m., An ego-motion algorithm based on the tracking of arbitrary curves, in *Proc. 2nd European Conf. Computer Vision*, Santa Margherita ligure, Italy, 1992. pp. 467-475.

[2] Cipolla, R. and Blake, A., Surface orientation and time to crash from image divergence and deformation, in *Proc. 2nd European Conf. Computer Vision*, Santa Margherita ligure, Italy, 1992. pp. 187-202.

[3] Flierl, M., Mavlankar, A., and Girod, B., Motion and disparity compensated coding for multiview video, *IEEE Transactions on Circuits and systems for Video Technology*, Vol. 17, No. 11, 1474-1484, 2007.

[4] Flierl, M. and Girod, B., Multiview video compression, *IEEE Signal Processing Magazine*, Vol. 24, No.6, 66-76, 2007.

[5] Girod, B., Aaron, A. M., Rane, S., and Rebollo-Monedero, D.; Distributed video coding, *Proceedings of the IEEE, Special issue on advances in video coding and delivery*, Vol. 93, No. 1, 71 - 83, 2005.

[6] Guo, X., Lu, Y., Wu, F., Gao, W., and Li, S., Distributed multi-view video coding, *Proceedings of SPIE*, San Jose, California, USA, Jan., 2006, Vol. 6077.

[7] Heikkilä, J., Geometric Camera Calibration Using Circular Control Points, *IEEE Trans. Pattern. Anal. Mach. Intell.*, Vol. 22, No. 10, 1066-1077, 2000.

[8] Hu, X. and Ersson, T., Active state estimation of nonlinear systems, *Automatica*, Vol. 40, No.12, 2075-2082, 2004.

[9] Jain, J. and Jain, A., Displacement measurement and its application interframe image coding, *IEEE Transaction on Communications.*, Vol. 29, No. 12, 1799-1808, 1981.

[10] Kang, Li-Wei and Lu, Chun-Shien, Multi-View Distributed Video Coding with Low-Complexity Inter-Sensor Communication Over Wireless Video Sensor Networks, *IEEE International Conference on Image Processing*, Sept. 16 2007-Oct. 19 2007, Volume 3, PP. 13 - 16.

[11] Levoy, M. and Whitted, T., The use of points as a display primitive, *UNC-Chapel Hill Computer Sceince Techinical Report*, Jan. 1985.

[12] Lukacs, M., Predictive coding of multi-viewpoint image sets, in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Tokyo, Apr. 1986, pp. 521-524.

[13] Maitre, M., Guillemot, C., and Morin, L., 3-D Model-Based Frame Interpolation for Distributed Video Coding of Static Scenes, *IEEE Transactions on Image Processing*, Vol. 16, No. 5, 1246 - 1257, 2007.

[14] Matveev, A., Hu, X., Frezza, and Rehbinder, H., Observer for systems with implicit output, *IEEE Trans. Aut. Contr.*, Vol. 45, No. 1, 168-173, 2000.

[15] Murray, R. M., Li, Z., and Sastry, S. S., *A Mathematical Introduction to Robotic manipulation*, CRC Press, Boca Raton, New York, 1994.

[16] Naemura, T., Tago, J., and Harashima, H., Real-time video-based modeling and rendering of 3D scenes, *IEEE Comput. Graph. Applicat.*, Vol. 22, No. 2, 66-73, 2002.

[17] Puri, R., Majumdar, A., Ishwar, P., and Ramchandran, K., Distributed video coding in wireless sensor networks, *IEEE Signal Processing Magazine*, Vol. 23, No. 4, 94 - 106, 2006.

[18] Slepian, J. D. and Wolf, J. K., Noiseless coding of correlated information source, *IEEE Transactions on Information Theory*, Vol. IT-19, 471-480, 1973.

[19] Soatto, S., Perona, P., Fressa, R., and Picci, P., Recursive motion and structure estimation with complete error characterization, vision,*Proc. IEEE Conf. Computer Vision Pattern Recognition.*, Pages 428-433, New York, June, 1993.

[20] Soatto, S., Fressa, R., and Perona, P., Motion estimation via dynamic vision, *IEEE Trans. Aut. Contr.*, Vol. 41, No. 3, 393-413, 1996.

[21] Taubin, G., Estimation of planar curves, surface and nonplanar space curves defined by implicit equations with applications to edge and range image segment, *IEEE Trans. Pattern. Anal. Mach. Intell.*, Vol. 13, No. 11, 1115-1138, 1991.

[22] Taylor, C. J. and Kriegman, D. J., Structure and motion from line segments in multiple images,*IEEE Trans. Pattern. Anal. Mach. Intell.*, Vol. 17, No. 11, 1021-1032, 1995.

[23] Toffetti, G., Tagliasacchi, M., Marcon, M., Sarti, A., Tubaro, S., and Ramchandran, K., Image compression in a multi-camera system based on a distributed source coding approach, *European Signal Processing Conference*, Antalya, Turkey, Sept., 2005.

[24] Wyner, A. D., Recent results in the Shannon theory, *IEEE Transactions on Information Theory*, Vol.IT-20., No. 1, 2-10, 1974.

[25] Wyner, A. D., On source coding with side infoermation at the decoder, *IEEE Transactions on Information Theory*, *IEEE Transactions on Information Theory*, Vol. IT-21, 294-300, 1975.

[26] Wyner, A. D. and Ziv, J., The rate-distortion function for source coding with side information at the decoder, *IEEE Transactions on Information Theory*, Vol. IT-22, 1-10, 1976.

[27] Wyner, A. D., The rate-distortion function for source coding with side information at the decoder, *IEEE Transactions on Information Theory*, Vol. IT-38, 60-80, 1978.
IEEE Transactions on Information Theory, *IEEE Transactions on Information Theory*, Vol. IT-21, 294-300, 1975.

[28] Zhang, Z. and Faugeras, O. D., Estimation of displacement from two 3d frames obtained from stereo, *IEEE Trans. Pattern. Anal. Mach. Intell.*, Vol. 14, No. 12, 1141-1156, 1992.

[29] Zhang, C. and Chen T., A self-reconfigurable camera array, *Eurographics Symposium on Rendering*, Norrkping, Sweden, Jun. 2004, pp. 243–254.

[30] Zhu, X., Aaron, A., Girod, B., Distributed compression for large camera arrays, *Proc. IEEE workshop on statistical signal processing*, St, Louis, Sept., 2003. PP. 30-33.