Bayesian Self-Calibration of a Moving Camera *

Gang Qian and Rama Chellappa

Center for Automation Research and
Department of Electrical and Computer Engineering
University of Maryland
College Park, MD 20742, U.S.A.
{gqian, rama}@cfar.umd.edu

Abstract. In this paper, a Bayesian self-calibration approach is proposed using sequential importance sampling (SIS). Given a set of feature correspondences tracked through an image sequence, the joint posterior distributions of both camera extrinsic and intrinsic parameters as well as the scene structure are approximated by a set of samples and their corresponding weights. The critical motion sequences are explicitly considered in the design of the algorithm. The probability of the existence of the critical motion sequence is inferred from the sample and weight set obtained from the SIS procedure. No initial guess for the calibration parameters is required. The proposed approach has been extensively tested on both synthetic and real image sequences and satisfactory performance has been observed.

1 Introduction

Automatic retrieval of the intrinsic parameters of a (relatively) moving camera from an observed image sequence has been of great interest to researchers in computer vision since the early 1990s. Subsequent to the pioneering work on camera self-calibration reported by Maybank and Faugeras [1,2], numerous algorithms have been proposed to calibrate cameras with constant (see [3] for a review) or varying intrinsic parameters [4,5,6,7]. Although significant efforts have been made to solve the self-calibration problem, several challenges still remain: sensitivity to observation noise, initialization of the algorithms, and processing of critical motion sequences (CMS) [8,9,10,11]. The first two challenges are common difficulties arising in nonlinear problems such as the camera self-calibration problem. In some situations, these two factors interact with each other and make the problem more complex. For example, large observation noise might create more local minima (or maxima) and can easily trap iterative optimization methods such as Levenberg-Marquardt and steepest-descent algorithms in local minima, preventing them from converging to the global optimal solution. Hence a good initial guess of the calibration parameters is needed.

* Partially supported by the U.S. Army Research Laboratory (ARL) Collaborative Technology Alliance contract DAAD19-01-2-0008.

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2351, pp. 277–293, 2002.
© Springer-Verlag Berlin Heidelberg 2002

In addition to the difficulties due to noise sensitivity and initialization, the existence of CMSs makes camera self-calibration even more difficult in practice. CMSs are sequences of camera motions resulting in inherent ambiguities in camera self-calibration and therefore ambiguities in uncalibrated Euclidean reconstruction [8]. Any practical self-calibration method must take into account the processing of CMSs since CMSs frequently occur in applications. Previous research [9] has shown that ambiguous Euclidean reconstructions from a CMS are conjugated and hypothesis verification can be used to detect and determine the type of CMS. Nevertheless, in the presence of noise, some camera motion sequences which are not CMSs can also result in ambiguous Euclidean reconstruction. Moreover, it has been recently reported in [12] that camera motion sequences "close" to CMSs in the sense of producing ambiguous Euclidean reconstructions can be far away from any type of CMSs in the motion sequence space in the sense of L_2 norm. If hypothesis verification is applied to this kind of sequence, it will be classified as one type of CMSs, and the true solution of the motion sequence, which is actually outside of CMSs, will be lost. Therefore, hypothesis verification is not sufficient in these circumstances.

In this paper, we focus on the main problem of self-calibration: estimation of the field of view (FOV) with all the other intrinsic parameters known. The unknown FOV can be either constant or varying throughout the image sequence. We develop a self-calibration algorithm, which is capable of processing CMS and yielding reasonable calibration estimates without any specific requirements of initialization. The new approach is developed based on the sequential importance sampling (SIS) technique. The SIS procedure is recently introduced by [13] to estimate the state parameters of a non-linear/non-Gaussian dynamic system. In SIS, the joint posterior distribution of the state parameters given the observations is approximated by a set of samples and their related weights. The SIS procedure has been used for solving the structure from motion (SfM) problem. An SIS-based SfM algorithm has been developed in [14] and it was shown to be robust to feature tracking errors and to be able to handle motion/structure ambiguities. However, in that case, all the intrinsic parameters of the camera are assumed to be given. In the paper, we still use the SIS method to attack the camera self-calibration problem because of its capability of solving problems involving non-linear systems.

2 Theoretical Background

2.1 Self-Calibration of a Moving Camera

In many practical situations, the calibration of the camera used to capture the sequences is not available, i.e. the intrinsic parameters of the camera such as the field of view (or the focal length relative to the film size), the position of principal point, skew factor and lens distortion are not known beforehand. To reconstruct accurate 3D Euclidean structure and motion, these intrinsic parameters have to be found.

Assume that a perspective projection camera model is considered and the lens distortion can be ignored or is already known, the following calibration

matrix is of interest:

$$\mathbf{A} = \begin{bmatrix} fk_u & fk_u \cot\theta & u_0 \\ 0 & \frac{fk_v}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where f is the focal length of the camera in world coordinate units. k_u and k_v are the lengths in pixels of the unit length of the world coordinate system in the vertical and horizontal directions, respectively. u_0 and v_0 are the pixel coordinates of the principal point in the image plane. θ is the angle between the vertical and horizontal axes in the image plane. Usually it is very close to $\pi/2$. Note that by writing the calibration matrix in the above form, we have moved the image plane to the front of the lens and have aligned the coordinate axes in the image plane with those in the world coordinate system.

A 3D point W has projection m in the image plane. Following the notation of Faugeras in [15], let $\mathbf{w} = [X, Y, Z]^T$ be the world coordinates of W and $\mathbf{m} = [u, v]^t$ be the pixel coordinates of its projection m. Let the homogeneous coordinates of any vector $\mathbf{v} = [v_1, v_2, \cdots, v_n]^T$ be $\tilde{\mathbf{v}} = [v_1, v_2, \cdots, v_n, 1]^T$. Hence, in the homogeneous coordinate system, we have $\tilde{\mathbf{w}} = [X, Y, Z, 1]^T$ and $\tilde{\mathbf{m}} = [u, v, 1]^T$. $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{w}}$ are related by a 3×4 projection matrix $\tilde{\mathbf{P}}$

$$\lambda\tilde{\mathbf{m}} = \tilde{\mathbf{P}}\tilde{\mathbf{w}} \tag{2}$$

where λ is called projective depth and does not play any role in the location of m in the image plane. Hence (2) is often rewritten as

$$\tilde{\mathbf{m}} \simeq \tilde{\mathbf{P}}\tilde{\mathbf{w}} \tag{3}$$

by ignoring λ, where the symbol $\simeq$ means that the two quantities are equal up to a scale factor. The projection matrix $\tilde{\mathbf{P}}$ can be decomposed as

$$\tilde{\mathbf{P}} = \mathbf{A}[\mathbf{R}| - \mathbf{R}\mathbf{t}] \tag{4}$$

where $\mathbf{A}$ is the calibration matrix and $(\mathbf{R}, \mathbf{t})$ is the displacement of the camera, containing both rotation and translation.

The problem of self-calibration is to estimate the calibration matrix $\mathbf{A}$ purely from an observed image sequence without any knowledge or control of the motion of the camera. In this paper, we will focus on the estimation of an unknown constant or varying FOV when all the other intrinsic parameters are given. We also assume that the camera moves continuously and takes many camera positions.

2.2 Critical Motion Sequences

In the research on solving the camera self-calibration problem, it has been observed that not all camera motion sequences lead to unique camera intrinsic parameters and 3D Euclidean scene reconstruction. Camera motion sequences that produce ambiguous calibrations are called *critical motion sequences*.

Identification of CMSs with various assumptions on the calibration has been systematically investigated in the literature. In [7], Sturm listed all CMSs when

the calibration parameters are constant. The CMSs for known calibration parameters, except for a varying FOV, can be found in [11]. Kahl also discussed CMSs when some intrinsic parameters can vary [10]. Other references on identification of CMSs can been found in [16,17,18]. Recall that we assume that the camera moves continuously and takes many camera positions. According to [10, 11], there are three types of CMSs for varying FOV that contain many camera positions:

- arbitrary translation with arbitrary rotation only about the optical axis
- translation along an ellipse or a hyperbola with the optical axis tangent to the ellipse or hyperbola
- translation along the optical axis with arbitrary rotation about the camera centers (at most two)

When the FOV is constant, only the first type of motion in the above CMS list is critical when many camera positions are present [8,11]. When the camera motion is not continuous, there are more critical motion sequences existing for the self-calibration of a camera with only an unknown FOV. Analyzing CMSs related to discontinuously moving cameras is beyond the scope this paper.

We mainly deal with the self-calibration ambiguities caused by the first kind of CMSs, since this kind of motion sequences are frequently encountered in practice. Because camera motion includes translation and rotation, the assumption of continuous camera motion implies that the rotation of the camera is also continuous if the camera rotates. Since the third type of CMSs contain at most two rotations about the camera centers, it rarely happens to a continuously rotating camera. Although the second type of CMSs is not explicitly considered in this paper, we have shown by experiments that it is possible to remove the self-calibration ambiguities caused by the type of CMSs if we assume that the 3D scene is rigid and non-planar.

To handle the first type of CMSs, we need to find out the transformations between true and false Euclidean reconstructions. Since the false Euclidean reconstruction is actually a projective reconstruction, it is different from the true Euclidean reconstruction by a projective transformation, $\mathbf{T}_\Phi$ [19].

Assume that a false Euclidean reconstruction has been found. Let $\Delta f^{(0)}$ be the ratio of the focal lengths of true and false Euclidean reconstructions in the initial time instant, i.e. $\Delta f^{(0)} = \frac{f_e^{(0)}}{f_\Phi^{(0)}}$ where $f_e^{(0)}$ and $f_\Phi^{(0)}$ are the true and false focal lengths at the initial time instant, respectively. Let $\mathbf{t}_\Phi$ and $\mathbf{t}_e$ be the translation vectors associated with the false and true reconstructions, respectively. Let (α_e, β_e) and $(\alpha_\Phi, \beta_\Phi)$ be the translation direction angles associated with $\mathbf{t}_e$ and $\mathbf{t}_\Phi$. The unit vector in the translation direction is given by $(\sin(\alpha)\cos(\beta), \sin(\alpha)\sin(\beta), \cos(\alpha))^T$. At any time instant, the true and false projection matrices are related by

$$\tilde{\mathbf{P}}_e = \tilde{\mathbf{P}}_\Phi \mathbf{T}_\Phi = \mathbf{A}_\Phi[\mathbf{R}| - \mathbf{R}\mathbf{t}_\Phi] \begin{bmatrix} \mathbf{T}_{11} & \mathbf{0}_3 \\ \mathbf{0}_3^T & \kappa \end{bmatrix} = \mathbf{A}_\Phi[\mathbf{R}\mathbf{T}_{11}| - \kappa\mathbf{R}\mathbf{t}_\Phi] \tag{5}$$

where

$$\mathbf{T}_{11} = \begin{bmatrix} \Delta f^{(0)} & 0 & 0 \\ 0 & \Delta f^{(0)} & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

and θ is the rotation angle about the optical axis. Hence $\mathbf{R}\mathbf{T}_{11} = \mathbf{T}_{11}\mathbf{R}$ and (5) can be written as

$$\tilde{\mathbf{P}}_e = A_\Phi[\mathbf{T}_{11}\mathbf{R}| - \kappa\mathbf{R}\mathbf{t}_\Phi] = A_\Phi\mathbf{T}_{11}[\mathbf{R}| - \mathbf{R}(\kappa\mathbf{R}^T\mathbf{T}_{11}^{-1}\mathbf{R}\mathbf{t}_\Phi)] \tag{7}$$

Hence,

$$\mathbf{t}_e = \kappa R^T\mathbf{T}_{11}^{-1}\mathbf{R}\mathbf{t}_\Phi = \kappa\mathbf{T}_{11}^{-1}\mathbf{t}_\Phi \tag{8}$$

After some straightforward algebra, we have

$$\alpha_e = \arccos\frac{\Delta f\cos\alpha_\Phi}{\sqrt{(\Delta f\cos\alpha_\Phi)^2 + \sin^2\alpha_\Phi}}\ ,\ \beta_e = \beta_\Phi \tag{9}$$

Furthermore, the resulting rotation angles are all the same for different ambiguous reconstructions since the rotation matrix does not change when the false focal length is replaced by the true one. If the focal length is free to vary, the relationship between the true and false focal lengths at time i is given by $f_e^{(i)} = \Delta f^{(0)} f_\Phi^{(i)}$.

Regarding the transformations among the 3D structures, we have the following relationship.

$$\tilde{\mathbf{w}}_e = \mathbf{T}_\Phi^{-1}\tilde{\mathbf{w}}_\Phi = \begin{bmatrix} \mathbf{T}_{11}^{-1} & \mathbf{0}_3 \\ \mathbf{0}_3^T & \kappa^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{S}_\Phi \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{T}_{11}^{-1}\mathbf{w}_\Phi \\ \kappa^{-1} \end{bmatrix} \text{ and } \mathbf{w}_e = \kappa\mathbf{T}_{11}^{-1}\mathbf{w}_\Phi \tag{10}$$

The transformations of motion and scene structure reconstructions reveal the relationships among different reconstructions. It will be employed in the design of a novel self-calibration algorithm in the next section.

2.3 Sequential Importance Sampling

The SIS method has been recently proposed for approximating the posterior distribution of the state parameters of a dynamic system [13]. Usually, the state space model of a dynamic system is described by observation and state equations. Denote the measurement by $\mathbf{y}_t$ and the state parameter by $\mathbf{x}_t$. The observation equation essentially provides $f_t(\mathbf{y}_t|\mathbf{x}_t)$, the conditional distribution of the observation given the state. Similarly, the state equation gives $q_t(\mathbf{x}_{t+1}|\mathbf{x}_t)$, the Markov transition distribution from time t to time $t+1$. Let $\mathcal{X}_t = \{\mathbf{x}_i\}_{i=1}^t$ and $\mathcal{Y}_t = \{\mathbf{y}_i\}_{i=1}^t$. Samples drawn from $\pi_t(\mathcal{X}_t) = P(\mathcal{X}_t|\mathcal{Y}_t)$, the posterior distribution of the states given all the available observations up to t, are needed to compute the ensemble statistics such as mean or modes. However, to directly draw samples from a complex, high-dimensional distribution is very difficult in practice. An alternative way to the approximation of the posterior distribution is by a set of samples called *properly weighted samples* and their corresponding weights [13].

Suppose $\{\mathcal{X}_t^{(j)}\}_{j=1}^m$ is a set of random samples properly weighted by the set of weights $\{w_t^{(j)}\}_{j=1}^m$ with respect to π_t and let g_{t+1} be a trial distribution. Then the recursive SIS procedure to obtain the samples and weights properly weighting π_{t+1} is as follows.

SIS steps: for $j = 1, \cdots, m$,

(A) Draw $X_{t+1} = \mathbf{x}_{t+1}^{(j)}$ from $g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})$. Attach $\mathbf{x}_{t+1}^{(j)}$ to form $\mathcal{X}_{t+1}^{(j)} = (\mathcal{X}_t^{(j)}, \mathbf{x}_{t+1}^{(j)})$.

(B) Compute the "incremental weight" u_{t+1} by

$$u_{t+1}^{(j)} = \frac{\pi_{t+1}(\mathcal{X}_{t+1}^{(j)})}{\pi_t(\mathcal{X}_t^{(j)})g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})}$$

and let $w_{t+1}^{(j)} = u_{t+1}^{(j)} w_t^{(j)}$.

It can be shown [13] that $\{\mathcal{X}_{t+1}^{(j)}, w_{t+1}^{(j)}\}_{j=1}^m$ is properly weighted with respect to π_{t+1}. Hence, the above SIS steps can be applied recursively to get the properly weighted set for any future time instant when the corresponding observations are available. The choice of the trial distribution g_{t+1} is very crucial in the SIS procedure since it directly affects the efficiency of the proposed SIS method. In our approach, we used $g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t) = q_{t+1}(\mathbf{x}_{t+1}|\mathbf{x}_t)$. It can be shown that in this case $u_{t+1} \propto f(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})$, which is the conditional probability density function of the observations at $t+1$ given the state sample $\mathbf{x}_{t+1}$ and it is also known as the likelihood function of $\mathbf{x}_{t+1}$ since the observations are fixed.

3 Bayesian Self-Calibration Using Sequential Importance Sampling

In this section, we design a camera self-calibration algorithm assuming that the camera has an unknown constant or varying focal length or equivalently FOV with all the other parameters given. Our goal is to find an algorithm that does not have any specific requirement for initialization and is able to detect and handle the CMSs. The performance of the algorithm should degrade gracefully as the noise level in the observations increases. SIS is used as the main computational framework because of its capability for solving problems involving non-linear systems.

3.1 Parameterization of the Camera Motion

Before discussing the parameterization of sensor motion, we introduce two 3D Euclidean coordinate systems used in our research. One coordinate system is attached to the camera and uses the center of projection of the camera as its origin. It is denoted by C. The Z axis of C is along the optical axis of the camera, with the positive half-axis in the camera looking direction. The X-Y plane of C is perpendicular to the Z axis with the X and Y axes parallel to the borders of the image plane. Also, the X-Y-Z axes of C satisfy the right-hand rule. The

other coordinate system is a world inertial frame, denoted by I. I is fixed on the ground. The coordinate axes of I are configured in such a way that initially, I and C coincide. When the camera moves, C travels with the camera and I stays at the initial position.

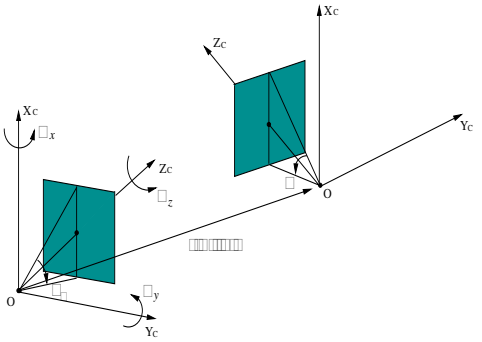

Fig. 1. Imaging model of a moving camera with an unknown FOV

Since the focal length (or FOV) is unknown, γ is used to represent the unknown focal length (or FOV). It is noted that the ranges of focal length and FOV are $[0, \infty]$ and $[0, \pi]$, respectively. Since a sampling based procedure is to be used, a naturally bounded variable is preferred. Hence, instead of focal length, the vertical FOV of the camera is to be estimated in the algorithm. Let γ denote the unknown FOV. Based on the above discussion, the state vector describing both extrinsic (motion) and intrinsic (FOV) parameters could be defined as

$$\mathbf{x} = (\psi_x, \psi_y, \psi_z, \alpha, \beta, \gamma). \tag{11}$$

Here (ψ_x, ψ_y, ψ_z) are the rotation angles of the camera about the coordinate axes of the inertial frame I and (α, β) are the elevation and azimuth angles of the camera translation direction, measured in the world system I. γ is the FOV of the camera. For simplicity, we still call $\mathbf{x}$, the motion parameter: remember that the FOV is now included in $\mathbf{x}$. If the FOV is free to change, one more component is added to the motion parameters.

$$\mathbf{x} = (\psi_x, \psi_y, \psi_z, \alpha, \beta, \gamma_0, \gamma) \tag{12}$$

where γ_0 represents the FOV of the camera at the initial time instant and γ denotes the FOV at other time instants.

State space model. Given the above motion parameterization, a state space model can be used to describe the behavior of a moving camera.

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{n}_x \text{ and } \mathbf{y}_t = Proj(\mathbf{x}_t, \mathcal{S}_t) + \mathbf{n}_y \tag{13}$$

where $\mathbf{x}_t$ is the state vector and $\mathbf{y}_t$ is the observation at time t. $Proj(\cdot)$ denotes the perspective projection, a function of camera motion $\mathbf{x}_t$ and scene structure

$\mathcal{S}_t$. n_x denotes the dynamic noise in the system, describing the time-varying property of the state vector. If no prior knowledge about motion is available, a random walk will be a suitable alternative for modeling the camera motion.

3.2 Processing Critical Motion Sequences

If a motion sequence is critical, the self-calibration algorithm should be able to detect the presence of CMSs. In our approach, CMS detection can be modeled as a hypothesis testing problem and the posterior probability of the non-criticalness of the motion sequence can be estimated. In the hypothesis testing problem, a binary variable $I_{\mathcal{C}}$ is introduced to indicate the presence of a CMS:

$$I_{\mathcal{C}} = \begin{cases} 1, \text{ the motion sequence is critical} \\ 0, \text{ the motion sequence is not critical} \end{cases} \tag{14}$$

This hypothesis testing problem can be naturally embedded in the SIS procedure. In the current case of interest, only one class of CMS exists: motion sequences that do not contain any rotation about an axis parallel to the image plane. Therefore, if the motion sequence is critical, motion samples with rotation only about the optical axis are enough to interpret the trajectories of the feature points. On the other hand, if the motion sequence is not critical, motion samples with rotation about axes parallel to the image plane have to be used to interpret the feature trajectories. Hence, two sets of samples are involved in the SIS procedure. The set of samples with only rotation about the optical axis is denoted by $\mathcal{X}_{\mathcal{C}}$ since it can explain the feature trajectories in the image plane introduced by the CMS. The other set of samples is denoted by $\mathcal{X}_{\mathcal{G}}$ because it will be used to explain the feature trajectories caused by general motion sequences other than CMSs. In the initialization stage of SIS, samples are generated in these two sets. Because no knowledge of the criticalness of the motion sequence is available at the beginning, equal numbers of samples are used in the two sets. The weights of the samples can be computed directly using the formula derived in [14].

During the motion of the camera, the criticalness of the motion sequence can change. A critical motion sequence up to time t can become non-critical at time $t+1$ if rotation about axes parallel to the image plane is present at time $t+1$. However, a non-critical motion sequence can never become critical. If the indicator $I_{\mathcal{C}}$ is viewed as the state of a dynamic system, this dynamic system can be characterized by a Markov chain. If the probability that a critical motion sequence becomes non-critical at time t is $P_{\mathcal{C}\to\mathcal{G}}(t)$, the state transition probabilities of the Markov chain are:

$$\begin{cases} P(I_{\mathcal{C}}(t+1)=0|I_{\mathcal{C}}(t)=1) = & P_{\mathcal{C}\to\mathcal{G}}(t) \\ P(I_{\mathcal{C}}(t+1)=1|I_{\mathcal{C}}(t)=1) = & 1-P_{\mathcal{C}\to\mathcal{G}}(t) \\ P(I_{\mathcal{C}}(t+1)=0|I_{\mathcal{C}}(t)=0) = & 1 \\ P(I_{\mathcal{C}}(t+1)=1|I_{\mathcal{C}}(t)=0) = & 0 \end{cases} \tag{15}$$

To take this fact into account in SIS, when drawing new samples for time $t+1$ from samples for time t, the samples in $\mathcal{X}_{\mathcal{C}}$ need to be transferred to $\mathcal{X}_{\mathcal{G}}$ with probability $P_{\mathcal{C}\to\mathcal{G}}(t)$. This can be done by adding rotation components about

axes parallel to the image plane, which can be drawn from a trial distribution. $P_{\mathcal{C}\to\mathcal{G}}(t)$ is unknown and no knowledge about it is available. Intuitively, 0.5 could be a good value for $P_{\mathcal{C}\to\mathcal{G}}(t)$ for all t since it gives the maximum uncertainty to the occurrence of the transformation of the motion sequence from critical to non-critical.

In the SIS procedure, the sample-weight set describes the posterior distribution of the motion parameters:

$$\begin{aligned} P(\mathcal{X}_t|\mathcal{Y}_t) &= \sum_{I_\mathcal{C}} P(\mathcal{X}_t, I_\mathcal{C}|\mathcal{Y}_t) \\ &= P(\mathcal{X}_t|I_\mathcal{C}=1,\mathcal{Y}_t)P(I_\mathcal{C}=1|\mathcal{Y}_t) + P(\mathcal{X}_t|I_\mathcal{C}=0,\mathcal{Y}_t)P(I_\mathcal{C}=0|\mathcal{Y}_t) \end{aligned}$$

The samples in $\mathcal{X}_\mathcal{G}$ are properly weighted by their corresponding weights with respect to $P(\mathcal{X}_t|I_\mathcal{C}=0,\mathcal{Y}_t)$, the posterior distribution of the motion parameters conditional on the motion sequence is not critical. The posterior probability of the presence of the critical motion sequence, $\pi_t(I_\mathcal{C}=1) = P(I_\mathcal{C}=1|\mathcal{Y}_t)$, can be obtained using the following theorem.

Theorem 1. *Assume that $\{\mathcal{X}_\mathcal{C}, \mathcal{X}_\mathcal{G}\}$ is properly weighted by $\{\mathcal{W}_\mathcal{C}, \mathcal{W}_\mathcal{G}\}$ with respect to $P(\mathcal{X}_t|\mathcal{Y}_t)$. $\mathcal{X}_\mathcal{C}$ is the sample set related to the hypothesis that a critical motion sequence is present and $\mathcal{W}_\mathcal{C}$ is the associated weight set. Then $\pi_t(I_\mathcal{C}=1)$, the posterior probability of criticalness of the given motion sequence, is given by*

$$\pi_t(I_\mathcal{C}=1) = \lim_{m\to\infty} \frac{\sum_{w_c\in\mathcal{W}_\mathcal{C}} w_c}{\sum_{w_c\in\mathcal{W}_\mathcal{C}} w_c + \sum_{w_g\in\mathcal{W}_\mathcal{G}} w_g} \tag{16}$$

where m is the number of samples.

The proof of the theorem is very straightforward and it is omitted due to the page limitation.

3.3 The Algorithm

Based on the discussion in the last section, a Bayesian camera self-calibration algorithm using SIS can be designed. Before the algorithm is presented, one more issue needs to be addressed.

Samples in $\mathcal{X}_\mathcal{C}$ are to be transferred to $\mathcal{X}_\mathcal{G}$ with a certain probability. Let the sample-weight pairs before this transfer be $\mathcal{S}_\mathcal{C} = (\mathcal{X}_\mathcal{C}, \mathcal{W}_\mathcal{C})$ and $\mathcal{S}_\mathcal{G} = (\mathcal{X}_\mathcal{G}, \mathcal{W}_\mathcal{G})$. The transfer is done by reducing the weights of the samples in $\mathcal{S}_\mathcal{C}$ to half of the original values such that the number of samples belonging to $\mathcal{S}_\mathcal{C}$ after resampling has been decreased by half . Then, the samples of $\mathcal{S}_\mathcal{C}$ are put into $\mathcal{S}_\mathcal{G}$ with the remaining half weights. Therefore, the new sample-weight sets after the transfer step become

$$\tilde{\mathcal{S}}_\mathcal{C} = (\mathcal{X}_\mathcal{C}, \frac{\mathcal{W}_\mathcal{C}}{2}) \ , \ \tilde{\mathcal{S}}_\mathcal{G} = (\{\mathcal{X}_\mathcal{C}, \mathcal{X}_\mathcal{G}\}, \{\frac{\mathcal{W}_\mathcal{C}}{2}, \mathcal{W}_\mathcal{G}\}) \tag{17}$$

After the transfer of samples, resampling is done to the samples in $\tilde{\mathcal{S}}_\mathcal{C}$ to prepare samples for the next time instant. For samples in $\tilde{\mathcal{S}}_\mathcal{G}$, two procedures will be

utilized. Resampling is applied to the samples from $\mathcal{S}_{\mathcal{G}}$. A crucial procedure, called *uniforming*, is applied to the samples transferred from $\mathcal{S}_{\mathcal{C}}$ to $\tilde{\mathcal{S}}_{\mathcal{G}}$. As we mentioned in the last section, due to the fact that only a finite number of samples are used to describe the posterior distribution of the parameters, the empirical distribution of the FOV is not uniform. Uniforming is used to explore the fact that the posterior distribution of the FOV should be uniform in $[0, \pi]$ if the motion sequence is critical no matter how the FOV is distributed according to the empirical samples and weights. Let $\mathcal{X}_{\mathcal{C}\to\mathcal{G}} = \{\mathbf{x}^{(j)}\}_{j=1}^{k}$ be the samples transferred with weights $\mathcal{W}_{\mathcal{C}\to\mathcal{G}} = \{\frac{w^{(j)}}{2}\}_{j=1}^{k}$ in $\tilde{\mathcal{S}}_{\mathcal{G}}$. These sample-weight pairs are denoted by $\mathcal{S}_{\mathcal{C}\to\mathcal{G}} = (\mathcal{X}_{\mathcal{C}\to\mathcal{G}}, \mathcal{W}_{\mathcal{C}\to\mathcal{G}}) = (\mathcal{X}_{\mathcal{C}}, \frac{\mathcal{W}_{\mathcal{C}}}{2})$. Assume that m samples are used in the SIS procedure. Uniforming applied to $\mathcal{S}_{\mathcal{C}\to\mathcal{G}}$ can be done in the following way:

<u>*Uniforming.*</u> For $j = 1, \cdots, k$,

(A) Uniformly draw $\Gamma = \{\gamma^{(i)}\}_{i=1}^{n_j}$, samples of FOV from $[0, \pi]$ where

$$n_j = \frac{m w^{(j)}}{2(\sum_{w_c \in \mathcal{W}_{\mathcal{C}}} w_c + \sum_{w_g \in \mathcal{W}_{\mathcal{G}}} w_g)}$$

For $i = 1, \cdots n_j$

(B) Compute the associated focal length in units of the height of the film: $f^{(i)} = (2 \tan \frac{\gamma^{(i)}}{2})^{-1}$. Motion sample $\mathbf{x}^{(j)}$ is used as a seed to produce more samples. Let the focal length associated with the FOV in $\mathbf{x}^{(j)}$ be $f^{(0)}$. By using the transformation among ambiguous motion estimates derived in section 2, the camera motion parameters related to the current focal length $f^{(i)}$ can be found directly using (9). A new motion sample can be formed as

$$\mathbf{x}_j^{(i)} = (0, 0, \Psi_z, \alpha_j^{(i)}, \beta, \gamma^{(i)}) \tag{18}$$

where Ψ_z and β are the corresponding components in $\mathbf{x}^{(j)}$. Hence, $\mathcal{X}_{\mathcal{U}}^{(j)} = \{\mathbf{x}_j^{(i)}\}_{i=1}^{n_j}$ are the new samples obtained from seed $\mathbf{x}^{(j)}$.

If the focal length is free to vary, the associated sample value of γ_0 needs to be changed properly. The new samples can be written as

$$\mathbf{x}_j^{(i)} = (0, 0, \Psi_z, \alpha_j^{(i)}, \beta, \gamma_0^{(i)}, \gamma^{(i)}) \tag{19}$$

where

$$\gamma_0^{(i)} = 2 \arctan \frac{f_0^{(0)} f^{(0)}}{f^{(i)}} \ , \ f_0^{(0)} = \left(2 \tan \frac{\gamma_0^{(0)}}{2}\right)^{-1} \tag{20}$$

and $\gamma_0^{(0)}$ is the value of FOV in the seed sample $\mathbf{x}^{(j)}$.

(C) $\{\mathcal{X}_{\mathcal{U}}^{(j)}\}_{j=1}^{k}$ contains the samples by uniforming the sample-weight pair $\mathcal{S}_{\mathcal{C}\to\mathcal{G}}$.

Based on the above discussion, the SIS procedure for Bayesian self-calibration proceeds as follows.

Bayesian Camera Self-Calibration Using SIS

1. **Initialization.** Draw samples of the motion parameters $\{\mathbf{x}_0^{(j)}\}_{j=1}^m$ from the initial distribution π_0. π_0 describes the distribution of the motion parameters $\mathbf{x}_0$ before the camera moves. The absence of camera motion does not imply that $\mathbf{x}_0 = 0$. Although the rotation angle vector ψ and the translational vector are zero, the translational angles can be uniformly distributed. Hence, in $\{\mathbf{x}_0^{(j)}\}$, the components of the rotation angles are all set to zero and the samples of α, β and γ (and γ_0 if the focal length is free to change) are drawn from the uniform distribution in $[0, \pi]$, $[0, 2\pi]$ and $[0, \pi]$, respectively. Since all the samples are drawn from the exact posterior distributions, equal weights are assigned to these samples. Since at the moment, rotation angles are all zeros, all the current samples belong to $\mathcal{X}_\mathcal{C}$ and $\mathcal{X}_\mathcal{G}$ contains no samples.
   For $t = 1, \cdots, \tau$:
2. **Sample transfer.** Two pairs of sample-weight sets are available: $(\mathcal{X}_\mathcal{C}, \mathcal{W}_\mathcal{C})$ and $(\mathcal{X}_\mathcal{G}, \mathcal{W}_\mathcal{G})$. Transfer all samples in $\mathcal{X}_\mathcal{C}$ to $\mathcal{X}_\mathcal{G}$, and assign half weight to each sample. Denote the sample-weight pair transferred from $\mathcal{S}_\mathcal{C}$ to $\mathcal{S}_\mathcal{G}$ by $\mathcal{S}_{\mathcal{C}\to\mathcal{G}} = (\mathcal{X}_\mathcal{C}, \frac{\mathcal{W}_\mathcal{C}}{2})$. The new sample-weight pairs after sample transfer are $\tilde{\mathcal{S}}_\mathcal{C} = (\mathcal{X}_\mathcal{C}, \frac{2\mathcal{W}_\mathcal{C}}{2})$ and $\tilde{\mathcal{S}}_\mathcal{G} = \{\mathcal{S}_{\mathcal{C}\to\mathcal{G}}, \mathcal{S}_\mathcal{G}\} = (\{\mathcal{X}_\mathcal{C}, \mathcal{X}_\mathcal{G}\}, \{\frac{\mathcal{W}_\mathcal{C}}{2}, \mathcal{W}_\mathcal{G}\})$.
3. **Resampling and uniforming.** Resample the samples in $\tilde{\mathcal{S}}_\mathcal{C}$ according to their associated weights. $\hat{\mathcal{X}}_\mathcal{C}$ is used to represent the set containing the resulting samples. For samples in $\tilde{\mathcal{S}}_\mathcal{G}$, uniforming and resampling are applied to samples belonging to different sets. Uniforming is performed on the samples in $\mathcal{S}_{\mathcal{C}\to\mathcal{G}}$. The samples originally in $\mathcal{X}_\mathcal{G}$ are then resampled. The sample set produced by these two procedures is denoted by $\hat{\mathcal{X}}_\mathcal{G}$. Since resampling and uniforming have been executed, all the samples in $\hat{\mathcal{X}}_\mathcal{C}$ and $\hat{\mathcal{X}}_\mathcal{G}$ have equal weights. Let $\{\hat{\mathbf{x}}_{t-1}^{(j)}\}_{j=1}^m$ denote the current samples.
4. **Sample generation.**
   For $j = 1, \cdots, m$:
   Draw $\mathbf{x}_t^{(j)}$ from the distributions of $\hat{\mathbf{x}}_{t-1}^j + \mathbf{n}_x$. The following distributions can be used for the dynamic noises in the translation direction angles. $n_\kappa \sim U(-\delta_\kappa, \delta_\kappa), \kappa \in \{\alpha, \beta\}$ where δ_α and δ_β can be chosen as positive numbers. The distributions of the dynamic noises in the rotation angles depend on $\hat{\mathbf{x}}_{t-1}^j$. If $\hat{\mathbf{x}}_{t-1}^j$ is in $\hat{\mathcal{X}}_\mathcal{C}$, disturbances are only added to the Z component of the rotation angles with $n_{\psi_z} \sim \mathcal{N}(0, \sigma_z)$. Otherwise, dynamic disturbances can be added to all the three components of the rotation angles and the associated distributions can be $n_{\psi_\iota} \sim \mathcal{N}(0, \sigma_\iota), \iota \in \{x, y, z\}$ where δ_x, δ_y and δ_z can also be chosen as some small positive numbers.
5. **Weight computation.** Compute the weights of the samples,$\{w_t^{(j)}\}$ using the weight computation formulas derived in [14] (See equations (5), (6) and (7) in [14] for details). Notice that in this case, the computation of the positions of terminal points of the epipolar line l involves not only the extrinsic parameters of the camera motion, but also γ, the field of view. The resulting samples and their corresponding weights $(\mathcal{X}_t^{(j)}, w_t^{(j)})$ are properly weighted with respect to $\pi_t(\mathcal{X}_t)$. If more image frames are available, go back to step 2.

Inference of Depth Distribution. By using the SIS procedure proposed above, the posterior distribution of the camera extrinsic and intrinsic parameters can be approximately described by the resulting samples and their corresponding weights. The inference of the posterior distribution of the depths, $\pi_t(\mathbf{z}_t)$, can be accomplished as follows. In [14], two algorithms are presented to find the posterior distribution of feature depths when all the camera calibration parameters are known. The discussion about the inference of $\pi_t(\mathbf{z}_t)$ based on the results obtained on $\pi_t(\mathcal{X}_t)$ is still valid here since nothing has changed except that the unknown FOV γ is included in the motion vector $\mathbf{x}$. Hence, the posterior distribution of the depths can be directly inferred using the samples and weights properly weighted with respect to the posterior motion (both extrinsic and intrinsic) distribution. Both algorithms developed in [14] can be used to find samples and weights properly weighted with respect to the posterior distribution of feature depths in this case except that the known constant value focal length used in [14] needs to be replaced by the values of FOV in motion samples.

4 Experimental Results and Performance Analysis

By using the proposed algorithm for Bayesian self-calibration, constant or varying FOV can be recovered and furthermore, the motion of the camera and the scene structure can be reconstructed.

4.1 Constant Field of View

Two experimental results using synthetic image sequences are presented first. The synthetic feature trajectories are corrupted by additive white Gaussian noise (AWGN). In the first experiment, the standard deviation (STD) of the AWGN is 0.5 pixel. We consider this case as a nominal case.

A Nominal Case Study. In this case, rotation about the X axis is present, hence the motion sequence is not critical.

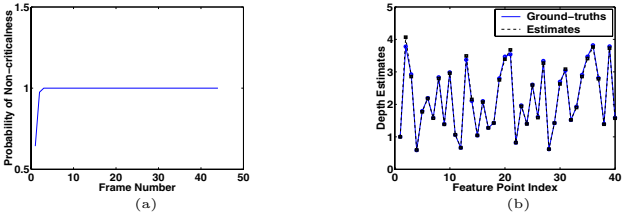

Fig. 2. (a) shows the probability of non-criticalness of the motion sequence in the nominal case and (b) shows the MMSE estimates of the feature depths.

Figure 2 (a) shows the probability of the non-criticalness of the motion sequence. The horizontal axis of Figure 2 (a) is the time axis and the corresponding value on the vertical axis indicates the probability of the non-criticalness of the

motion sequence up to that time. It can be seen that this probability starts with a relatively low value at the beginning of the sequence. The reason is as follows. At the beginning of the sequence, the rotation about the X axis is small. Due to the observation noise in feature correspondences, the sequence looks like a critical motion sequence. Along with the increase in the rotation angle about the X axis, the probability of non-criticalness of the sequence approaches 1 eventually. This indicates that this motion sequence is not critical.

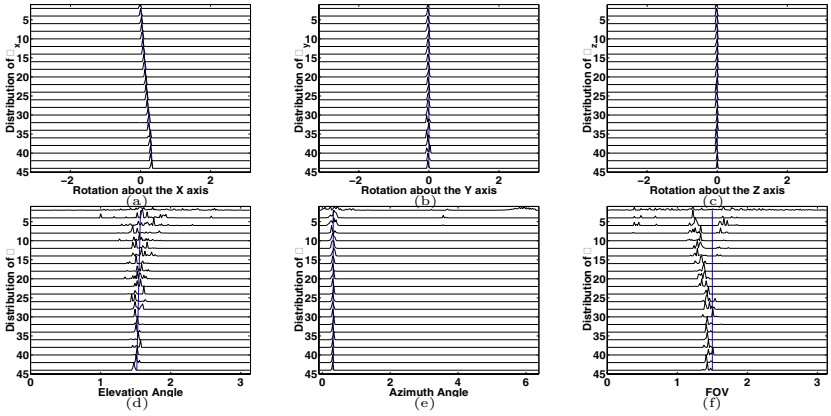

Fig. 3. Camera motion and calibration distributions in the nominal case

The motion distributions are shown in Figure 3. The ground-truths (including FOV) are indicated by the bold solid lines in Figure 3. Figures (a,b,c) show the distributions of the rotational angles ψ_x, ψ_y and ψ_z, respectively. The following two figures give the distributions of the translational angles α and β. In each figure, the distribution of the corresponding motion parameter at each time instant is shown from the top of the figure to the bottom. (ψ_x, ψ_y, ψ_z) are in the range $[-\pi, \pi]$. α is in $[0, \pi]$ and β in $[0, 2\pi]$. All the other motion distribution results in this paper can be interpreted in the same way. We can see that the resulting posterior distributions of motion parameters have peaks very close to the ground-truths. Figure 3(f) shows the ground-truths and the posterior distributions of the field of view. Figure 2 (b) shows the ground-truths and the minimum mean square error (MMSE) estimates of the depths of the feature points. Since they are very close, it is difficult to distinguish one from the other.

A Critical Motion Sequence. Critical motion sequences were also generated to test the proposed algorithm. One example is included here. In this example, the virtual camera only translates along the horizontal axis without any rotation.

Figure 4 (a) shows the probability of non-criticalness of this motion sequence. It can be seen that the probability of non-criticalness of the camera stays at zero throughout the sequence, indicating that this motion sequence is critical. The ground-truth and posterior distribution of the FOV are shown in Figure 4 (b) and it can be seen that the FOV is nearly uniformly distributed.

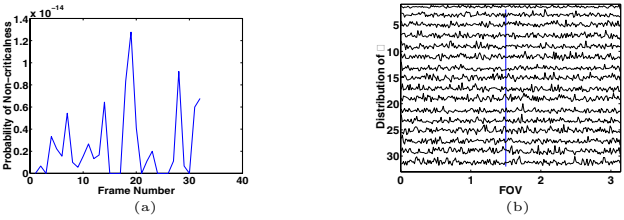

Fig. 4. (a) shows the posterior probability of non-criticalness of the motion sequence. Since in the experiment the motion is pure translation, it can be seen that this probability is very close to zero, indicating that the motion sequence is critical. and (b) is the distribution of the FOV and it can be seen that the FOV is nearly uniformly distributed.

4.2 Freely Varying Field of View

Now let us look at examples when the FOV of the camera can freely vary.

Circular Camera Motion. We let the virtual camera move along a circle. At the same time, the field of view of the camera is enlarged.

Figure 5 (a) - (g) shows the ground-truths and posterior distributions of the motion parameters. Figure 5 (h) shows the probability of non-criticalness of the motion sequence. It can be seen that this probability increases to 1 when more image frames are used and the rotation angles about the X axis increases. The MMSE estimates of depths are shown in Figure 5 (i).

Elliptical Camera Motion. In this example, we tested the proposed algorithm using an image sequence produced by a virtual camera moving along an ellipse with the optical axis of the camera tangent to the ellipse. Recall that this type of motion sequence was found critical [10,11], when the FOV can varying. The feature points are spread randomly in the 3D space and they are not on a plane. The feature correspondences are corrupted by AWGN with one-pixel of STD. By using the proposed approach, the posterior motion distribution of the camera can be approximated by a set of samples and their weights.

Figure 6 shows the refined estimates after applying the Levenberg-Marquardt non-linear optimization, using the result from the SIS algorithm as an initial guess. In Figure 6, the horizontal axis of each plot is the time axis. The solid lines show the ground-truths of the motion and calibration parameters of the camera and the dashed lines are the estimates of the parameters. It can be seen that the final results are very close to the ground-truths. Hence, it has been experimentally shown that it is possible to remove the calibration ambiguity introduced by the motion along an ellipse, which is the second type of CMS mentioned in Section 2.2.

3D Face Modeling Using Uncalibrated Camera. In this example, an image sequence containing 17 frames was captured using SunCamera II, which is an adjustable CCD color camera. The vertical FOV of SunCamera II is 33 degrees,

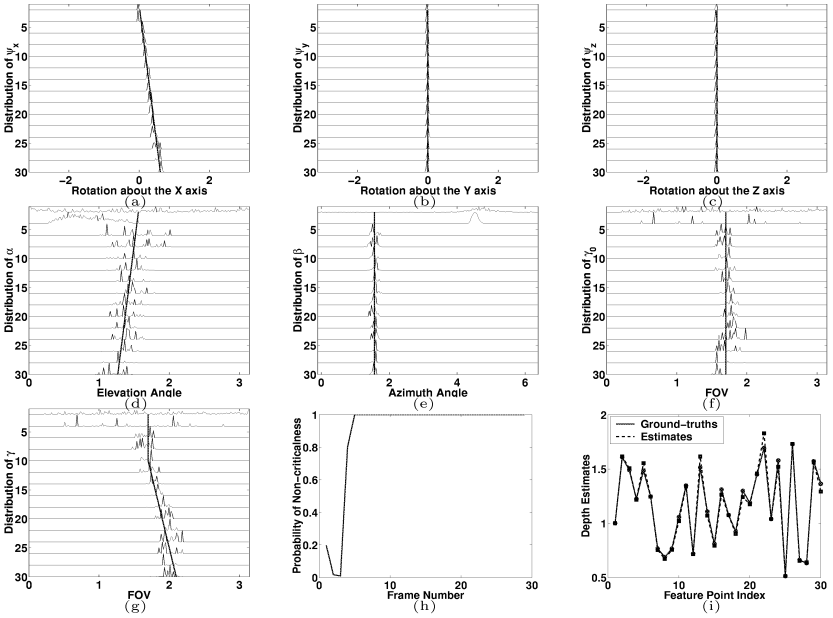

Fig. 5. The ground-truths and the posterior distributions of the camera motion and calibration parameters in the case of circular motion with varying focal length. (a)-(c) are the plots for camera rotation angles. (d) and (e) are the plots for camera translation direction angles. (f) is for the FOV of the camera at the initial time instance. (g) is for the varying FOV at different time instances. (h) shows the probability of non-criticalness of the motion sequence and (i) is the MMSE estimate of the feature depths.

which is equal to 0.576 radian. By using the proposed Bayesian self-calibration algorithm, the FOV of the camera can be accurately estimated and a 3D face model can be reconstructed. The MMSE estimate of the vertical fov is 0.5804 radian, which is very close to the ground-truth.

In Figure 7, (a) shows the texture map of this face sequence and (b)-(d) are the reconstructed 3D face model viewed from different angles.

5 Conclusions

In this paper, we have presented an algorithm for camera self-calibration using SIS. Our efforts have concentrated on the main problem of self-calibration: estimation of the FOV with all the other intrinsic parameters known, where the unknown FOV can be either constant or varying throughout the image sequence. The proposed algorithm is capable of processing the CMSs and quasi-CMSs and it does not have any specific requirements for initialization. The proposed method has also been tested extensively and satisfactory experimental results are obtained. A future research direction could be the extension of the algorithm

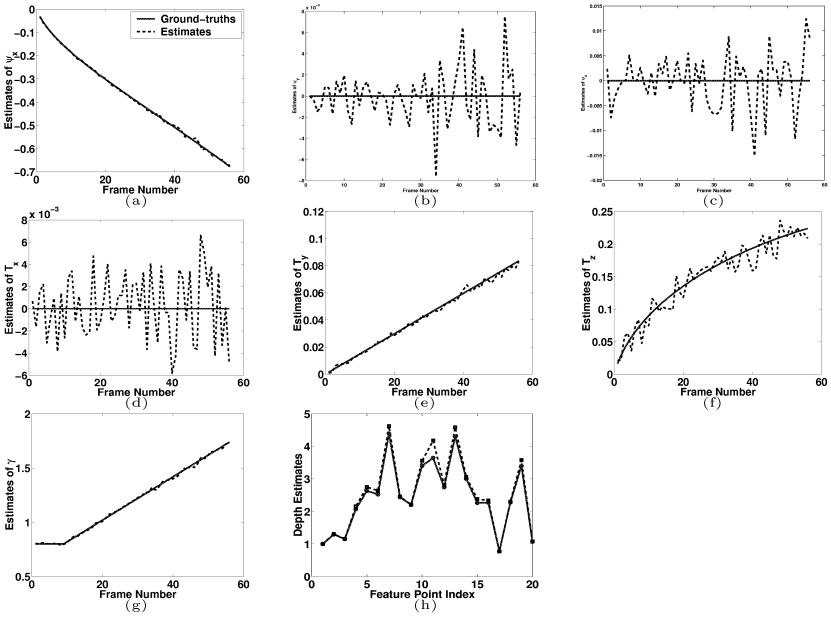

Fig. 6. The ground-truths and the estimates of the camera motion, calibration and structure parameters after non-linear optimization in the case of elliptical motion. (a)-(c) are the plots for camera rotation angles. (d)-(f) are the plots for camera translations. (g) shows ground-truth and estimate of the field of view of the camera. (h) is depth estimate.

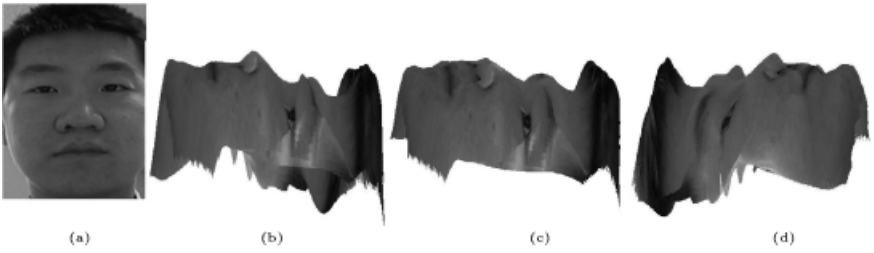

Fig. 7. The texture map and reconstructed 3D model from an uncalibrated face sequence

to self-calibration with more unknown camera intrinsic parameters such as the position of the principle point and the aspect ratio.

References

1. Faugeras, O., Luong, Q., Maybank, S.: Camera self-calibration: Theory and experiments. In: European Conference on Computer Vision, Santa Margherita Ligure, Italy. (1992) 321–334

2. Maybank, S., Faugeras, O.: A theory of self-calibration of a moving camera. International Journal of Computer Vision **8** (1992) 123–151
3. Fusiello, A.: Uncalibrated Euclidean reconstruction: a review. Image and Vision Computing **18** (2000) 555–563
4. Enciso, R., Vieville, T.: Self-calibration from four views with possibly varying intrinsic parameters. Image and Vision Computing **15** (1997) 293–305
5. Pollefeys, M., Koch, R., van Gool, L.: Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. International Journal of Computer Vision **32** (1999) 7–25
6. de Agapito, L., Hayman, E., Reid, I.: Self-calibration of a rotating camera with varying intrinsic parameters. In: British Machine Vision Conference, Southampton, UK. (1998)
7. Kahl, F., Heyden, A.: Euclidean reconstruction and auto-calibration from continuous motion. In: International Conference on Computer Vision, Vancouver, Canada. (2001) II: 572–577
8. Sturm, P.: Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition, San Juan, PR. (1997) 1100–1105
9. Sturm, P.: Critical motion sequences and conjugacy of ambiguous Euclidean reconstructions. In: Scandinavian Conference on Image Analysis, Lappeenranta, Finland. (1997)
10. Kahl, F., Triggs, B., Astrom, K.: Critical motions for auto-calibration when some intrinsic parameters can vary. Journal of Mathematical Imaging and Vision **13** (2000) 131–146
11. Sturm, P.: Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. In: British Machine Vision Conference, Nottingham, UK. (1999) 63–72
12. Pollefeys, M., Van Gool, L.: Do ambiguous reconstructions always give ambiguous images? In: International Conference on Computer Vision, Vancouver, Canada. (2001) II: 187–192
13. Liu, J.S., Chen, R.: Sequential Monte Carlo methods for dynamic systems. J. Amer. Statist. Assoc. **93** (1998) 1032–1044
14. Qian, G., Chellappa, R.: Structure from motion using sequential Monte Carlo methods. In: International Conference on Computer Vision, Vancouver, Canada. (2001) II: 614–621
15. Faugeras, O.: Three-Dimensional Computer Vision: A Geometric Viewpoint. MIT Press (1993)
16. Kahl, F., Triggs, B.: Critical motions in Euclidean structure from motion. In: IEEE Computer Vision and Pattern Recognition, Fort Collins, CO,. (1999) II:366–372
17. Kahl, F.: Critical motions and ambiguous Euclidean reconstructions in auto-calibration. In: International Conference on Computer Vision, Corfu, Greece. (1999) 469–475
18. Kahl, F., Astrom, K.: Ambiguous configurations for the 1D structure and motion problem. In: International Conference on Computer Vision, Vancouver, Canada. (2001) I: 184–189
19. Hartley, R.: Projective reconstruction and invariants from multiple images. PAMI **16** (1994) 1036–1041