Recursive 3D Model Reconstruction Based on Kalman Filtering

Ying-Kin YU, Kin-Hong WONG and Michael Ming-Yuen CHANG

Abstract—**A recursive two-step method to recover structure and motion from image sequences based on Kalman filtering is described in this paper. The algorithm consists of two major steps. The first step is an extended Kalman filter for the estimation of the object's pose. The second step is a set of extended Kalman filters, one for each model point, for the refinement of the positions of the model features in the 3D space. These two steps alternate from frames to frames. The initial model converges to the final structure as the image sequence is scanned sequentially. The performance of the algorithm is demonstrated with both synthetic data and real world objects. Analytical and empirical comparisons are made among our approach, the interleaved bundle adjustment method and the Kalman filtering based recursive algorithm by Azarbayejani and Pentland. Our approach outperformed the other two algorithms in terms of computation speed without loss in the quality of model reconstruction.**

Index Terms: **3D structure acquisition, Structure from motion, Kalman filtering, Multimedia processing**

I. INTRODUCTION

THE research work presented in this paper falls into the category of structure from motion in the field of computer vision. The goal of these kinds of researches is to reconstruct a 3D structure and its pose from a sequence of 2D images. A subset of the problems is to perform the recovery of structure and pose in a sequential manner. This is also known as recursive or casual structure from motion. There are many real world applications for the casual model reconstructions. One novel application is that the pose of the camera in a movie scene recovered can be used to produce augmented reality video, in which synthetic objects can be mixed with real world characters. Another application is to use these algorithms to produce 3D movies. In that, the 3D scene is reconstructed from the frames of

Manuscript received July 31, 2003. This work was supported by the Research Grant Council of Hong Kong Special Administrative Region. (Project No. CUHK4389/99E)

Y.K.Yu and K.H.Wong are with the Department of Computer Science and Engineering, the Chinese University of Hong Kong, Hong Kong. (e-mail: ykyu@cse.cuhk.edu.hk, khwong@cse.cuhk.edu.hk).
M.M.Y.Chang is with the Department of Information Engineering, the Chinese University of Hong Kong, Hong Kong. (e-mail: mchang@ie.cuhk.edu.hk).

the movies sequentially. The viewers are allowed to change the viewpoints of the scenes as the movie is played.

There are various techniques to deal with the 3D reconstruction problem. One of the most popular approaches is the use of epipolar geometry. With the known correspondences between the two views, a constraint between these views can be set up. The camera motion can be recovered from the Fundamental matrix up to a scale factor if the camera is fully calibrated. The structure of the 3D model can be found be solving a set of equations. In a similar sense, the technique has been extended to three views or more [7].

Factorization [5] [6] is another common approach to tackle the problem of structure from motion. The work in [5] demonstrates the approach under the assumption of orthographic projection. The factorization method has been extended to handle the reconstruction of multiple independent moving objects [6]. Bundle adjustment is also an effective method [9]. It minimizes the re-projection error between the estimated model and the image measurements. The minimization procedure can be done in batch either by the well-known Newton's or Levenberg-Marquardt iteration. A branch of it is the interleaved bundle adjustment as described in [2] and [9]. It breaks up the minimization problem into two steps so as to reduce the size of the Jacobian involved, resulting in speeding up the algorithm.

The methods mentioned previously tackle the problem in a batch, in which the structure and motion are optimized for all the images at one time. There are solutions that recover the structure and motion in a sequential way. Most of them are based on Kalman filtering. The work in [3] finds the pose of the object based on a known CAD model from stereo images in real-time. The results are applied for visual servoing of robot manipulators. Some researchers adopt iterated extended Kalman filter (IEKF) for updating the structure in Euclidean [4] or projective framework [1]. The pose and the structure of the object are recovered alternately by a RANSAC-based equation solving technique and the IEKF. Thomas and Oliensis apply standard Kalman filtering to fuse the recent structure estimates found by Horn's algorithm using the most recent image pair with the previous structure estimates at each time-step [12].

The series of methods in [10] [11] [13] [14] recover both the structure and motion in a recursive manner. The work by Broida et al [14] is the ancestor of this series of researches. They apply a single IEKF to recover the structure and pose of the object. Azarbayejani and Pentland describe a method in [11] that makes significant improvements over [14]. Extension is made to recover the focal length of the camera in addition to the pose and structure. The pointwise structure is represented by one parameter per point such that the computation is overdetermined at every frame when the number of features is larger than 7, resulting in a better convergence and stability of the filter. The most recent

work of recursive structure recovery is by Chiuso et al [10]. Similar techniques in structure from motion have also been applied to simultaneous localization and map-building for robot navigations [16].

The two-step Kalman filter based algorithm presented in this paper is inspired by the methods of interleaved bundle adjustment as described in [9]. The main advantage of our two-step approach is that we achieve a linear time and space complexity in terms of the available model features. This saves a lot of computation when the number of features needed to be handled is large, which is quite common for the reconstruction of objects with full details. In addition, our implementation can handle the structure from motion problem with changeable set of feature points. The full 360° view of an object can be reconstructed, which is demonstrated in the experiment.

The rest of this paper is organized as follows. The modeling of our problem is first introduced in Section II. In section III, the overview of the two-step algorithm is described. In section IV and V, the formulations of the EKF for pose estimation and structure updating are presented. In section VI, the handling of the changeable set of feature points in our implementation is discussed. In section VII, there is an analytical comparison among our Kalman filter based approach, the interleaved bundle adjustment method [2], and the recursive algorithm by Azarbayejani and Pentland [11]. In section VIII, some experiments with real and synthetic data are performed. The results from the three approaches mentioned are analyzed.

II. Problem Modeling

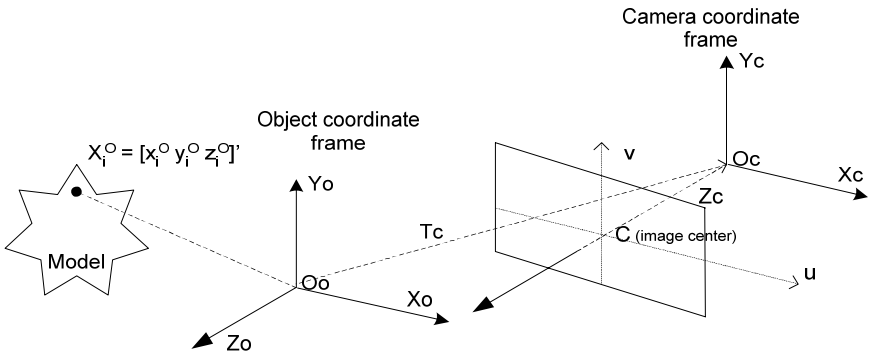

Figure 1. The geometric model of the system

Figure 1 describes the geometry of the model reconstruction system. $X_i^O = [x_i^O, y_i^O, z_i^O]$ and $X_i^C = [x_i^C, y_i^C, z_i^C]$ denote the coordinates of the point *Xi* with respect to the object and the camera coordinate frame respectively. A point on

the image plane is denoted by $p_i = [u_i, v_i]$. The relationship between the object frame and the camera frame is as follows:

$$X_i^C = (RX_i^O + T) + T^C \tag{1}$$

R is a 3x3 rotation matrix and T is a 3x1 translation vector. T^C is a 3x1 translation vector that brings the object in the object frame to the camera frame. It is a constant in the model recovery process. Camera used in the system is calibrated. It has a fixed focal length f. The camera model is full perspective. The problem of structure from motion in our system is to recover the coordinates of model point X_i^O in the object coordinate frame and the pose of the object, i.e. the rotation R and translation T, with respect to the views at each time-step.

III. Overview of the Algorithm

The model reconstruction system is divided into three parts: feature extraction and tracking, model initialization, structure and pose updating. To make the presentation of the algorithm easy to understand, we first assume that all the features are observable from the first to the last frame in the image sequence. The details of the extra treatments needed to handle the changeable set of feature points are discussed in section VI.

A. Feature extraction and tracking

The KLT tracker described in [8] is used to extract feature points and track them from images to images. In our work, it is assumed that the problem of tracking has been solved and point matches from KLT are reliable enough for model reconstruction.

B. Model initialization

The model initialization is achieved by assuming that the projection of the first image in the sequence is orthographic. Perspective projection is assumed for image formation process in the remaining frames. The orthographic projection is expressed mathematically as:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} = \frac{f}{z_{init}} \begin{bmatrix} x_i^C \\ y_i^C \end{bmatrix} \tag{2}$$

z_{init} is the distance between the object and camera center. It is a parameter given by the user of the system and can be approximated easily. To obtain the initial model, features in the first image are back-projected from the image plane to the camera coordinate frame according to equation (2). The resulting initial structure is a planar model

located at a distance z_{init} from the camera.

C. *Structure and pose updating*

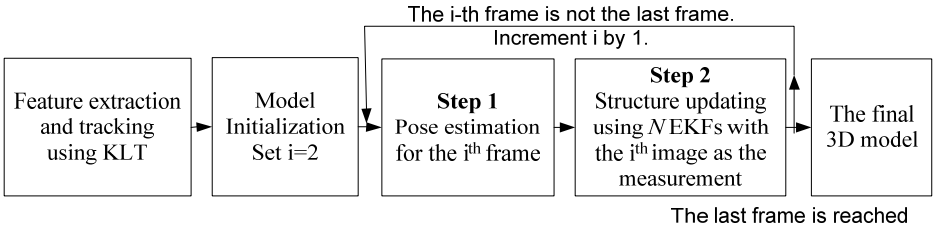

Figure 2. The flowchart of our two-step Kalman filter based algorithm.

The initial model and the second image are fed into the first step for pose estimation. An EKF is adopted. The pose of the object with respect to the second image is estimated. The newly recovered pose and the input image are passed to the second step of the algorithm for structure updating. The second step consists of a set of N EKFs. Each filter corresponds to one coordinate point in the reconstructed 3D model. With the observations and the pose recovered for the current image frame, the coordinates of each feature point are updated accordingly. The algorithm alternates between the step 1 and 2 until all images in the sequence are used.

IV. Structure updating

The following is the formulation of the EKF for structure updating. For N model points, N EKFs are needed for the structure update. For simplicity, we consider only one point in the model denoted by X_t, where t is the time-step of the model point, with respect to the object coordinate frame. $\hat{X}_{t,t-1}, \hat{X}_{t,t}$ are the positions of point X_t after the prediction and update respectively.

We first define the dynamic model of a 3D point. The state transition and observation equation for a model point can be written as:

$$X_t = X_{t-1} + \gamma_t$$

$$\varepsilon_t = h_t(X) + v_t$$

γ_t and v_t are the zero mean Gaussian noise with covariance Q_t and R_t respectively. ε_t is the real measurement from the image sequence. $h_t(X)$ is the projection function of the system:

$$h_t(X) = f\left[\frac{x^C}{z^C} \quad \frac{y^C}{z^C}\right]^T$$

X^C is obtained from equation (1) by substituting X^O by X. The EKF first provides an optimal estimate of the state at the next sample time in accordance with the following equations:

$$\hat{X}_{t,t-1} = \hat{X}_{t-1,t-1}$$
$$\Lambda_{t,t-1} = \Lambda_{t-1,t-1} + Q_t$$

They are known as the prediction equations. Λ is the 3x3 covariance matrix of $\hat{X}$. Here a noise covariance Q_t of γ is added. In normal situations, only the entry corresponds to the z coordinates of the 3D point is set to non-zero. The reason is that the initial guess of the structure is a planar object. Followed by the state prediction, the filter improves the previous estimate using the measurements acquired:

$$\hat{X}_{t,t} = \hat{X}_{t,t-1} + W(\varepsilon_t - h_t(\hat{X}_{t,t-1}))$$
$$\Lambda_{t,t} = \Lambda_{t,t-1} + W\nabla h_X \Lambda_{t,t-1}$$

$$W = \Lambda_{t,t-1}\nabla h_X^T(\nabla h_X \Lambda_{t,t-1}\nabla h_X^T + R_t)^{-1}$$

They are known as the update equations. W is the 3x2 Kalman gain matrix of the filter. R_t is the measurement noise covariance matrix. It is a tuning parameter and is set according to the quality of the images. It can also be acquired during the process of camera calibration. ∇h_X is the Jacobian of the non-linear projection function $h_t(X)$ evaluated at $\hat{X}_{t,t-1}$. In this way, the coordinates of the model points can be updated accordingly.

V. Pose Estimation

The dynamic model that describes the motion of the object is as follows. w is the state of the system and is defined as:

$$w = \left[t_x \quad \dot{t}_x \quad t_y \quad \dot{t}_y \quad t_z \quad \dot{t}_z \quad \alpha \quad \dot{\alpha} \quad \beta \quad \dot{\beta} \quad \gamma \quad \dot{\gamma}\right]$$

t_x, t_y, and t_z are the translations of the object along the x, y and the z axes respectively. $\dot{t}_x, \dot{t}_y, \dot{t}_z$ are their corresponding velocities. α, β, γ are the Yaw, Pitch and Roll angles. Their corresponding angular velocities are $\dot{\alpha}, \dot{\beta}, \dot{\gamma}$. T_s denotes the duration over the sample period. Over the sample period, the velocities are assumed constant. The state transition and observation equation for the model are:

$$\hat{w}_t = A\hat{w}_{t-1} + \gamma'_t,\ A = diag\left\{\begin{bmatrix}1 & T_s\\ 0 & 1\end{bmatrix},..,\begin{bmatrix}1 & T_s\\ 0 & 1\end{bmatrix}\right\}$$

$$\varepsilon'_t = g_t(w_t) + v'_t$$

γ'_t and v'_t are the zero mean Gaussian noise with covariance Q_t' and R_t' respectively. A is a 12x12 block known as the state transition matrix. ε'_t is a mx1 column vector representing the image measurements for m selected feature points. $g_t(w_t)$ is the projection function:

$$g_t(w_t) = f\begin{bmatrix}\frac{x_1^C}{z_1^C} & \frac{y_1^C}{z_1^C} & \cdots & \frac{x_i^C}{z_i^C} & \frac{y_i^C}{z_i^C} & \cdots & \frac{x_n^C}{z_n^C} & \frac{y_n^C}{z_n^C}\end{bmatrix}^T$$

Similar to $h_t(X)$, X^C is obtained from the equation (1) by substituting X^O by X. The rotation matrix R and translation matrix T are evaluated with the parameters of the column vector w_t.

In our system, a fixed number of feature points (e.g. 150 in our experiment) extracted by the tracker are passed to the EKF for pose estimation. The model points are chosen based on how much they are updated in the step of structure refinement. Those points that are steady and have a high tendency to remain at the same 3D coordinates are used. The reason is that less update on a point implies that the point is in an accurate position.

Here are the four core Kalman filter equations for pose estimation. The prediction equations for calculating the optimal estimates are:

$$\hat{w}_{t,t-1} = A\hat{w}_{t-1,t-1}$$
$$P_{t,t-1} = AP_{t-1,t-1}A^T + Q_t'$$

The update equations for the corrections of estimates are:

$$\hat{w}_{t,t} = \hat{w}_{t,t-1} + K(\varepsilon'_t - g_t(\hat{w}_{t,t-1}))$$
$$P_{t,t} = P_{t,t-1} - K\nabla g_w P_{t,t-1}$$

$$K = P_{t,t-1}\nabla g_w{}^T(\nabla g_w P_{t,t-1}\nabla g_w^T + R'_t)^{-1}$$

$\hat{w}_{t,t-1}$ and $\hat{w}_{t,t}$ are the states of w_t after the prediction and update respectively. $P_{t,t-1}$ and $P_{t,t}$ are 12x12 matrices that correspond to the covariances of $\hat{w}_{t,t-1}$ and $\hat{w}_{t,t}$. K is the 12x2m Kalman gain matrix. ∇g_w is the Jacobian of the non-linear observation equation $g_t(w)$ evaluated at $\hat{w}_{t,t-1}$. In this way, the pose of the model to the next frame is estimated.

VI. Handling the changeable set of feature points

The set of "active" feature points is changing due to occlusion and disocclusion. Extra treatments are needed in each step of the model reconstruction process. In feature tracking, the whole image sequence is divided into a number of sections. For each section, a number of frames near the end of that section are overlapped with the frames at the beginning of the succeeding section. The KLT tracker in [8], with feature replacement mechanism, is applied to each section independently. Cutting the image sequence into sections forces the tracker to release obsolete features after a finite time limit. It is important because wrong point correspondences arise without this treatment.

New model points in the structure are initialized when new point features appear in the image sequence. This is obtained by assuming the projection of that point on its first appeared image frame is orthographic. The initial position, expressed in the camera coordinate frame, is computed according to equation (2). The coordinates are then transformed back to the object coordinate frame by equation (1). After the initialization of the 3D position, a new EKF is set up for updating its position. In addition, this new point is added to the pool ready to be selected for pose estimation. No modification is needed in the EKF for pose estimation.

When a point feature vanishes from the image sequence, the filter that corresponds to the point is removed. The 3D position of that feature will no longer be updated. The index of that feature is also marked invalid for pose estimation since no related measurements in future time-steps can be used for finding the pose of the object. The treatments for handling changeable feature set are simple in our algorithm compared to the procedure in [10]. No sub-filters are required in our approach.

VII. Analytical Comparison with other algorithms

A. The interleaved bundle adjustment method

1) Computation efficiency

The main advantage of our Kalman filter based recursive approach is the gain in speed and scalability. An extra view of the object can be handled naturally by calculating the prediction and update equations for both the pose and structure only for that new measurement. However, the interleaved bundle adjustment method needs to re-compute from the first frame to the latest frame for a several iterations. Application like 3D movie as described in the introduction requires our recursive approach unless the film is pre-processed using a high performance computer.

2) Convergence of solutions

Another advantage of our Kalman filter based method is that it has a better convergence rate in handling long sequences with large object motion, say a total of 90 degrees rotation along one of the axes. Consider the rotation of the object between the first and the last frame. In the first pass and the first iteration of the interleaved bundle adjustment method, the initial model is used to estimate the pose of the object with respect the last frame. This results in a large pose estimation error and eventually leads the computation in the second pass of the interleaved bundle adjustment to diverge. For our recursive approach, the model is updated incrementally from frames to frames, the most up-to-date model is used for pose estimation. The problem of using an inaccurate model to compute the pose is eliminated.

B. The EKF by Azarbayejani and Pentland

1) Algorithm Complexity

The major difference between our approach and the EKF in [11] is that the structure refinement and pose estimation is broken down into two steps and each correspondence point in the structure is decoupled. In our approach, there are one 12x1 state vector for pose estimation plus N 3x1 state vectors for structure updating, where N is the total number of available features. This respectively results in one 12x12 and N 3x3 state covariance matrices. Since the number of measurements involved in the pose estimation step is fixed, both the storage and computation complexity are $O(N)$. For the EKF described in [11], the pose and structure are encoded in a single state vector. The size of the state covariance matrices is $(N+7)$x$(N+7)$. The storage and computation complexity are $O(N^2)$ and $O(N^3)$ respectively. Our modification is actually a tradeoff between speed and accuracy. However, experimental results show that the loss in accuracy is little and acceptable in real applications. Moreover, the reduction in complexity is useful for real-time robotics application since the computation resources in microcontrollers are tight. Also, our algorithm is ready to be implemented on distributed micro-processing system. The set of N EKFs used for structure updating can be run in parallel in a set of N micro-processors in a robotic system to speed up the computation.

2) The problem of scaling

Another advantage of our algorithm is that the exact model can be reconstructed given the alignment of translation T^C between the object coordinate frame and camera coordinate frame. The EKF by Azarbayejani and Pentland is subject to scaling problem even T^C is given. This is due to the nature of the structure model adopted in the filter. A

small deviation in the estimation of the Z coordinates of the point features causes a significant change in the scale of the object. The problem can be fixed by giving the EKF the real 3D coordinates of one of the feature points and setting their corresponding entries in the state covariance matrix zero.

VIII. Experiments and results

A. *Experiments with synthetic data*

The first set of experiments was conducted with synthetic data. A synthetic object with 300 random feature points in 3D within a cube of volume of $0.13m^3$, centered at a place 0.33m away from the viewing camera, was generated. The camera has a focal length of 6mm. It imposes a 2D zero mean Gaussian noise with standard deviation 0.1 pixels on the image captured. The object was moving with a steady motion at a rate of [0.005 0.01 0.02] degrees and [0.001 0.002 0.0003] meters for *[Yaw Pitch Row]* and *[Tx Ty Tz]* respectively. Random noise of 0.01 degrees was added to each rotation angle and a noise of 0.0005 meters was added to each of the translation parameter. 300 frames were generated for each test and a total of ten independent tests were carried out. Our Kalman filter algorithm, the interleaved bundle adjustment method [2] and the EKF by Azarbayejani and Pentland [11] were tested with the data and their results were compared.

Figure 3-5 show the results. The solid line, dotted line and the dash line are for our Kalman filtering approach, the interleaved bundle adjustment method and the EKF by Azarbayejani and Pentland respectively. The implementations of the three algorithms are in Matlab with a Pentium III 1GHz machine. Figure 3 demonstrates the convergence of 3D model error versus CPU time. Here the model error is defined as the percentage of mean-square-error of the 3D coordinates in the object coordinate frame. In our approach, the 3D model converges quickly to the solution within the first ten frames of the sequence. The average error is 0.69%. For the interleaved bundle adjustment method, the accuracy of the resulting model falls to 0.33% after 50 iterations. But for fair comparison, we should concentrate on the results after the first iteration. In this case, the average error is 0.74%, which is similar to our approach. The result is reasonable since our recursive algorithm cannot optimize the structure error of the model for all the images in the sequence simultaneously or optimize the error for scanning the sequence more than once, resulting in a larger final error than the batch processing approach. This is one of the characteristics of all recursive reconstruction algorithms. The plot of 3D model error of the EKF by Azarbayejani and Pentland is not included here. The reason is that the

model recovered by their algorithm is subject to scaling problem as described in section VII B 2). What can be concluded is that our Kalman filtering approach can find an accurate estimate of the structure in a shorter period of time than the interleaved bundle adjustment method.

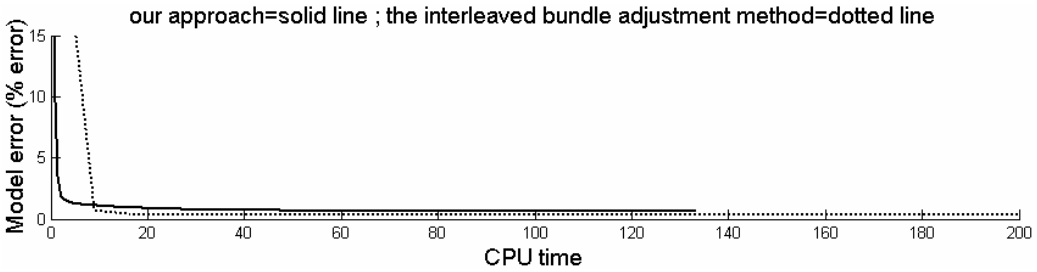

Figure 3. The 3D model error versus CPU time (in seconds) elapsed for the two approaches.

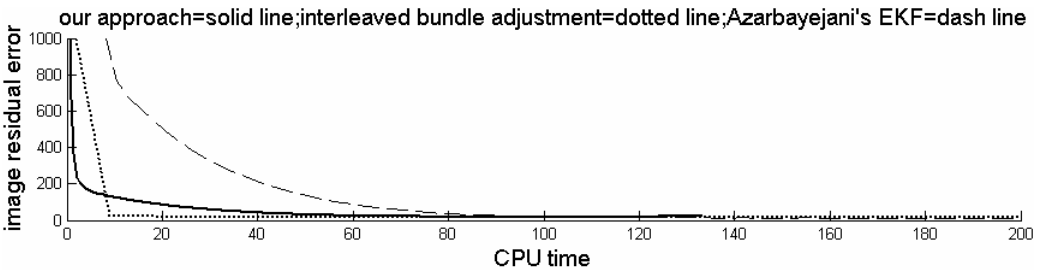

Figure 4. A graph showing the relationship between the CPU time and the image residual error among the three algorithms.

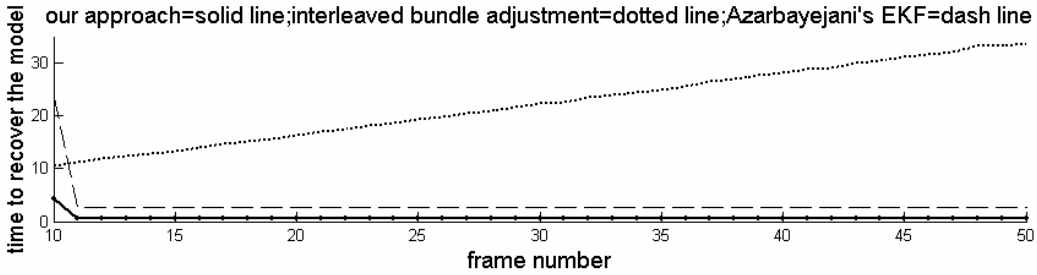

Figure 5. A graph showing the time needed to reconstruct a model when extra frames are added to the image sequence. Note that the interleaved bundle adjustment algorithm is set to run 20 iterations in the model reconstruction process.

Figure 4 shows the time for the three algorithms to optimize the image residual error of the back-projected model. By careful analysis, our algorithm minimizes the residual error to a low level in the shortest time among the three algorithms. Our approach finishes the processing of the 300-frame sequence in 133 seconds. The EKF by

Azarbayejani and Pentland and the interleaved bundle adjustment method complete at 786 and 436 seconds respectively. The figure also indicates that our approach falls to an error similar to the EKF by Azarbayejani and Pentland, which reveals that there is no significant loss in the structure acquisition accuracy even the filter is decoupled.

Figure 5 shows the time needed to reconstruct a model when extra frames were added sequentially to the image sequence. The first step in this plot was to reconstruct a model with the first 10 frames. The succeeding 40 frames were sequentially fed to the algorithm as the new measurements of the scene. Our approach outperformed the other two algorithms. It takes only 0.5 seconds to update the structure of the scene for every extra frame added to the image sequence. The EKF by Azarbayejani and Pentland takes about 3 seconds while the interleaved bundle adjustment method needs at least 10 seconds to do the same task.

B. Experiments with real scene

Experiments using real scene images were performed. Two test image sequences, one for the reconstruction of a paper box and one for a house model, were used in the experiment. The objects were put on a rotating turntable. Images were taken with a commercial web camera. The lengths of the paper box and the house sequence are 200 and 80 frames respectively.

Our recursive model reconstruction algorithm was applied to acquire the 3D models. After that, the wire-frame of the objects was built and texture from the appropriate images in the sequence was mapped to the recovered structure. The resultant objects were output in the form of VRML files. Figure 6 shows the results of the experiment.

We have successfully reconstructed the 360° view of the paper box model from its 2D images. The total number of point features present in the model is about 500. The quality is good in general. However, you may notice that there is a presence of outlying model features in the recovered structure, resulting in flaws in some parts of the model. The outlying model features are mainly due to point mismatches that arise from the process of feature tracking by the KLT tracker. For the recovered house model, the total number of model features in the scene is about 1000. More geometric details are revealed in this case. This demonstrates that our algorithm is able to handle complex scene reconstruction within a reasonable time limit.

IX. Conclusions

A two-step Kalman filter based algorithm is proposed and implemented. Our algorithm achieves linear time and space complexity in terms of the number of available point features. Our algorithm needs respectively less than one-forth and one-third the time of the EKF by Azarbayejani and Pentland and the interleaved bundle adjustment method to process a sequence with 300 point features per frame. Besides, an elegant method is proposed to handle the changeable set of point features. The method is applied to reconstruct the full 360^{o} view of a paper box in the real data experiment.

The major limitation of our model reconstruction system is caused by the feature tracker and the texture mapping method. Feature tracking and texture mapping play a significant role in the reconstruction of 3D models from the real images. One possible improvement to our system is to make a filter on the top of the original KLT tracker to eliminate the outlying point features in neighboring images with the fundamental matrix [15]. Other improvements, like incorporating a robust mechanism for choosing the texture from images, can be made to improve the quality of model reconstruction.

X. Acknowledgment

The work described in this paper was supported by a grant from the Research Grant Council of Hong Kong Special Administrative Region. (Project No. CUHK4389/99E)

References

[1] P.A. Beardsley, A.Zisserman and D.W.Murray, “Sequential updating of projecting and affine structure from motion”, Intl. Journal of Computer Vision 23, pp235-259, 1997.

[2] M.M.Y.Chang and K.H.Wong, “Model and pose acquisition using extended Lowe’s method”, IEEE Trans. on Multimedia (accepted)

[3] V.Lippiello, B.Siciliano and L.Villani, ‘Objects motion estimation via BSP tree modeling and Kalman filtering of stereo images”, Proc. of the IEEE Intl. Conf. on Robotics and Automation Washington DC, pp. 2968-2973, 2002.

[4] C.G.Harris and J.M.Pike. “3D positional integration from image sequence”, Image and Vision Computing Vol 6 No.2, 1988.

[5] C.Tomasi and T.Kanade, “Shape and motion from image streams under orthography: A factorization method”, Intl. Journal of Computer Vision 9(2), 137-154, 1992.

[6] J.Costeira and T.Kanade, "A Multibody Factorization method for independently moving objects", Intl. Journal of Computer Vision 29(3), 159-179, 1998.

[7] R. Hartley and A Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2002.

[8] C.Tomasi and T.Kanade, "Detection and Tracking of Point Features", Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

[9] B.Triggs, P.McLauchlan, R.Hartley and A.Fitzgibbon, "Bundle adjustment –A modern synthesis" In proc. of the Intl. Workshop on Visual Algorithm: Theory and Practice. pp 298-372, Corfu Greece, 1999.

[10] A.Chiuso, P.Favaro, H.Jin and S.Soatto, "Structure from motion casually integrated over time", IEEE Trans. on PAMI, vol24, No. 4, 2002.

[11] A.Azarbayejani and A.P.Pentland, "Recursive estimation of motion, structure, and focal length", IEEE Trans. on PAMI, vol 17, no 6, June 1995.

[12] J. Inigo Thomas and J.Oliensis, "Recursive multi-frame structure from motion incorporating motion error", Proc. DARPA Image Understanding Workshop, 1992.

[13] J.Weng, N.Ahuja and T.S.Huang, "Optimal motion and structure estimation", IEEE Trans. on PAMI, vol 15, no. 9, September 1993.

[14] T.J.Broida, S.Chandrasiiekhar and R.Chellappa, "Recursive 3-D motion estimation from monocular iamge sequence", IEEE Trans. on Aerospace and Electronic Systems, vol 26, no. 4, July 1990.

[15] S. Gibson, J. Cook, T. L. J. Howard, R. J. Hubbold, and D. Oram. "Accurate camera calibration for off-line, video-based augmented reality". In IEEE and ACM ISMAR 2002, Sep 2002. Darmstadt, Germany.

[16] A.J.Davison and D.W.Murray, "Simultaneous localization and map-building using active vision", IEEE Trans. on PAMI, Vol.24, No.7, July 2002.

Mr. Ying-Kin Yu received a B.Eng with first class honours in Computer Engineering from the Chinese University of Hong Kong in 2002. He is now a graduate student in the Computer Science and Engineering Department in the same university. He has been awarded the Sir Edward Youde Memorial Fellowship twice for his academic achievements. His research interests are computer vision, augmented reality and genetic algorithms. His contact address is: The Computer Science and Engineering Dept., The Chinese University of Hong Kong, Shatin, Hong Kong. Email: ykyu@cse.cuhk.edu.hk

Prof. Kin-Hong Wong received a B.Sc. in Electronics and Computer Engineering from the University of Birmingham in 1982, and a Ph.D. from the Engineering Dept. of the University of Cambridge, U.K. in 1986. He was a Croucher research fellow

at the University of Cambridge from 1985 to 1986. Prof. Wong joined the Computer Science Dept. of CUHK in 1988 and is now an Associate Professor. His research interests are 3D computer vision, virtual reality image processing, pattern recognition, microcomputer applications and computer music. His contact address is: The Computer Science and Engineering Dept., The Chinese University of Hong Kong, Shatin, Hong Kong. Email: khwong@cse.cuhk.edu.hk

Prof. Michael Ming-Yuen Chang received the B.Sc. in electrical engineering from Imperial College, London University and the PhD degree in electrical engineering from University of Cambridge in 1988. He then joined the Department of Information Engineering, The Chinese University of Hong Kong and is now an Associate Professor. His current research interest is in character recognition, scientific visualization and intelligent instrumental control. His contact address is: The Information Engineering Dept., The Chinese University of Hong Kong, Shatin, Hong Kong. Email: mchang@ie.cuhk.edu.hk

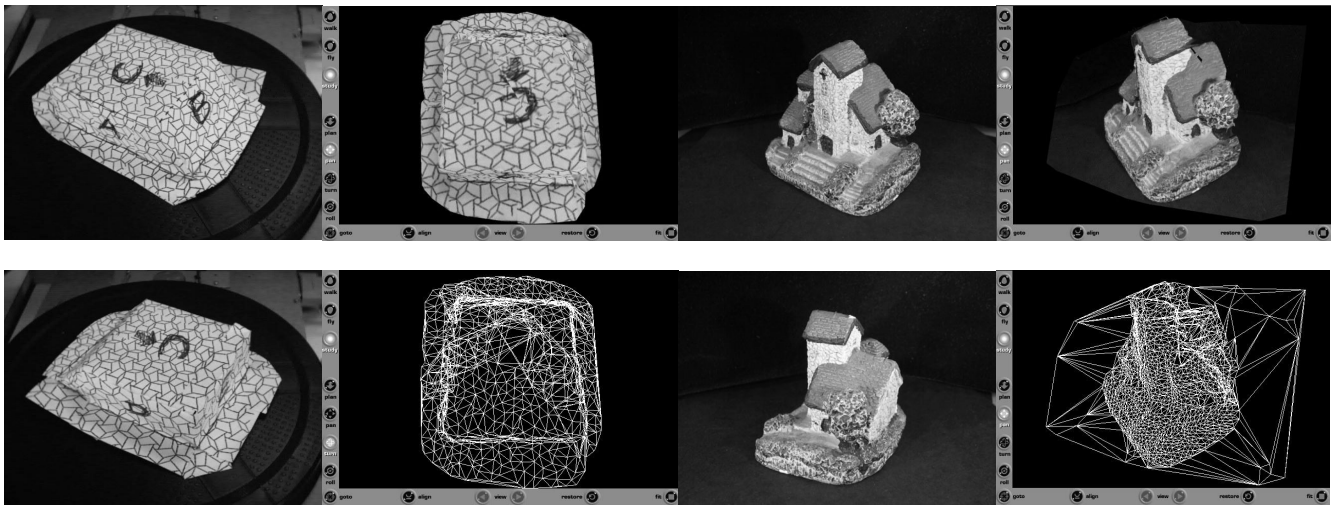

Figure 6. The results of model reconstruction. First column: The first and the 100th image of the paper box sequence. Second column: The reconstructed 3D paper box model viewed in Cortona. One view with texture mapping (the top one) and its wire-frames (the bottom one). Third column: The first and the last (80th) image of the house model sequence. Forth column: The reconstructed 3D house model. One view with texture mapping (the top one) and its wire-frames (the bottom one). More results can be found at http://www.cse.cuhk.edu.hk/~khwong/demo/