A KALMAN FILTERING APPROACH TO 3-D IR SCENE PREDICTION USING SINGLE-CAMERA RANGE VIDEO

Mehmet Celenk, James Graham, Don Venable, and Mark Smearcheck
School of Electrical Engineering and Computer Science
Ohio University, Stocker Center, Athens, OH 45701, USA

ABSTRACT

This paper presents a Kalman filtering approach to predicting 3-D video infrared (IR) scenes as a CMOS multi-coordinate axis sensory-camera mounted on a mobile vehicle moves forward in a controlled environment. Potential applications of this research can be found in indoor/outdoor heat-change based range measurement, synthetic IR scene generation, rescue missions, and autonomous navigation. Experimental results reported herein dictate that linear Kalman filtering based scene prediction accurately estimates future frames in range and intensity sensing. The low least mean square error (LMSE), on the average of 1%, proves the reliability of the approach to IR scene prediction. Currently, the proposed method is devised for piece-wise linear motion of the sensory system as it navigates in hall-way or corridor.

Keywords: 3D scene, range images, scene prediction, Kalman filtering, linear camera motion

1. INTRODUCTION

Recently, 3-D scene modeling has gained importance as imaging devices have acquired high-resolution range sensing capability. This, in turn, has lead to a great research effort in the vision area toward developing more accurate and reliable solutions to such diverse problems as unmanned navigation and guidance, surveillance, tracking, mapping, virtual world simulation, precision manufacturing, multimedia networking, animation, and rescue missions. Aside from the consumer-end applications, indoor and outdoor navigation has been greatly influenced from imaging-science developments and newly introduced range measurement devices. To this end, many 3-D image-motion estimation and modeling problems have been widely investigated in various indoor and outdoor navigation settings [1], [2], [3], [6], [9], [10], [11], [15], [19]. Determining egomotion (i.e., finding the motion of the camera based on its output image in response to the scene in question) has also been studied in the research effort discussed in [4]. Vosselman and Dijkman [5] presented a scheme to create 3-D regions via point clouds and planar region growing. Rozovskii and Petrov [7] successfully presented an optimal nonlinear filtering method for tracking-before-detection in IR image sequences. Another method that applies range based imagery was developed by King, et al. [8]. Their scheme makes use of multiple range scans for recognizing locations. Kim and Kweon [12] presented a robust and effective feature map integration method for IR target recognition. They reported that noise in an IR image causes estimation and shape matching instabilities in a target recognition system. Infrared imaging itself has also been extensively studied. Ibarra-Castanedo et al.'s work [13] discusses the difficulties in analyzing infrared data with a focus toward non-destructive testing. In [14], the problems of detecting point targets in IR images are examined. The research effort summarized herein has shown that it is relatively difficult to detect, extract, or highlight features in IR images. Hence, an approach that blindly searches for such attributes would likely produce false positives. A Kalman-based prediction, however, could yield image estimates more immune to noise. This, in turn, introduces a premise that instead of creating a model to match with the scene measurements, as is done in much of the research on object detection and tracking [3, 6], the scene estimates guide the tracker or classifier. The research presented in this paper uses an adaptive linear estimator on time varying 3-D scene data to predict future scenes. After modeling the frame-to-frame positions and velocities of each pixel, a Kalman filter is implemented for finding the future scene state. The effectiveness of this approach is tested by comparing the predicted scene to the future scene sensed using the Swiss Ranger SR 3000 CMOS camera [17, 18].

2. DESCRIPTION OF OVERALL APPROACH

The scene prediction method discussed in this work originated with the premise that IR scene modeling difficulty can be eased by accompanying the process with scene estimation and prediction. The main advantage of not attempting to model and observe the features of a 3-D scene is the alleviation of unreliable feature extraction and the promotion of a robust adaptive system that adjusts itself as the environment changes. The Kalman filtering approach to 3-D scene prediction is comparable to estimation and prediction, in concept. The simplest method that one can design is to assume that the next state of a moving object will be in the same direction as the previous trajectory. This simple assumption holds true for piecewise linear-wise motion. Of course, it is often supplemented by several other computationally complex approaches. Figure 1 is an example of the type of data that is used, and shows the x, y, z, intensity, and resulting point cloud image of a frame taken from the SwissRanger Camera [17]. In this work, future frames are predicted using the current and previous image frames. There are many approaches that tackle the motion of overlapped objects in a scene; however, the approach in this paper does not take this problem into account without losing its generality since the idea is not to detect motion, but to predict the scene with respect to the egomotion of the camera.

The system is designed to predict future frame contents by way of stochastic estimation and Kalman filtering. The future frame information is predicted with the current frame information. Next, the prediction error (e) is checked against the chosen threshold, and then the Kalman filter parameters are adjusted as necessary in order to meet the threshold. The predicted values, $\left[\hat{x}^- \quad \hat{y}^- \quad \hat{z}^-\right]^T$, are compared with the actual measured values, $\left[\hat{x} \quad \hat{y} \quad \hat{z}\right]^T$, and the resulting error is fed to the Kalman filter to be used for the next iteration.

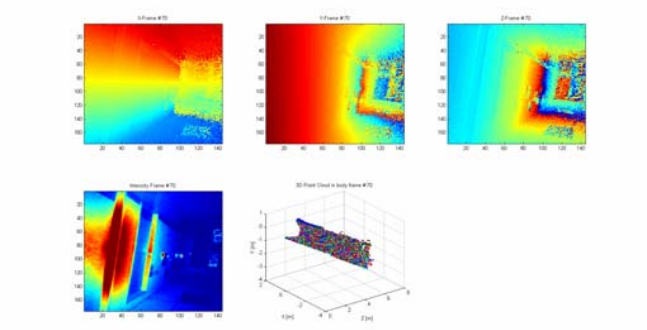

Fig.1. SwissRanger (x,y,z), intensity, and point-cloud images of a hallway.

The data is internally represented as 3 independent range images, each of which corresponds with the value in that particular Cartesian direction. Furthermore, as the scene changes (either due to the position / orientation changes of the camera, or from a moving object) the scene will move/rotate with a certain velocity. In this study it is assumed that the sensing unit moves (or accumulates data) slowly enough that changes in the scene can be observed linearly. Three images of a frame are sensed through x, y, and z coordinate axes allowing Kalman prediction to be performed on each axis independent of the others.

The Kalman estimator designed in this study relies on a linear prediction model that can be expressed as

$$\begin{bmatrix} x_{k+1} & y_{k+1} & z_{k+1} \end{bmatrix}^T = A \cdot \begin{bmatrix} x_k & y_k & z_k \end{bmatrix}^T + (t_{k+1} - t_k) \cdot \begin{bmatrix} \dot{x}_k & \dot{y}_k & \dot{z}_k \end{bmatrix}^T \quad (1)$$

where $\mathbf{A}$ is a rotation and scaling matrix, (x_k, y_k, z_k) are the current range values in each axis, and $(\dot{x}_k, \dot{y}_k, \dot{z}_k)$ are the current velocity estimates, and T is the matrix transposition operation. The indices k and $k+1$ represent the current and future state values. For computational simplicity, the filter operates on $m \times n$ windows from each axis of the full image. A linear Kalman filter's equations may be written as [16]

$$\mathbf{x_{k+1}} = \phi_k \mathbf{x_k} + \mathbf{w_k} \quad (2)$$

$$\mathbf{w_k} = \left[\ldots, pn_{(i-1),j}, pn_{i,j}, pn_{(i+1),j}, \ldots, vn_{(i-1),j}, vn_{i,j}, vn_{(i+1),j}, \ldots \right]^T \quad (3)$$

$$\mathbf{Q_k} = \mathbf{w_k} \mathbf{w_k}^T \quad (4)$$

where the state vector x_k will be of length $2 \cdot m \cdot n$ in the form

$$\mathbf{x}_k = \left[\ldots, x_{(i-1),j}, x_{i,j}, x_{(i+1),j}, \ldots, \dot{x}_{(i-1),j}, \dot{x}_{i,j}, \dot{x}_{(i+1),j}, \ldots \right]^T$$

These values are taken from pixels inside the m by n window of the scene that is currently being operated on as shown in Figure 2. In the current state, x_{ij} represents the pixel value at point i,j in the local $m \times n$ window, while $\dot{x}_{ij}$ denotes the respective velocity of the pixel at the designated point in each of the axial frames. In Eqn. 2, $\mathbf{x_{k+1}}$ is the propagated state vector one time epoch ahead of the current state. The state transition matrix ϕ_k is the Kalman filter implementation of matrix $\boldsymbol{A}$ of Eqn.1.

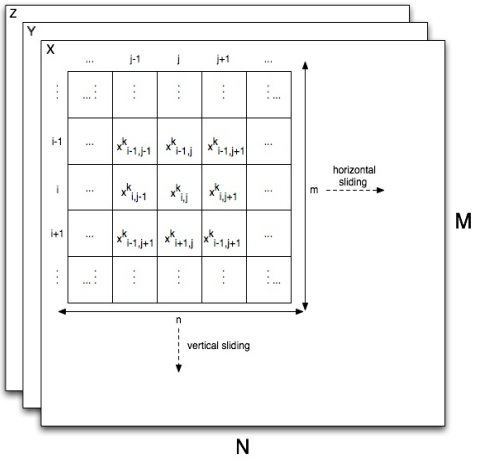

Fig. 2. m *x* n local window operating on (x, y, z) images.

The resulting state transition matrix equates to

$$\phi_k = \begin{bmatrix} I & I \\ 0 & I \end{bmatrix} \quad (5)$$

where, I is an $m \times n$ sized identity matrix and 0 is an $m \times n$ sized zero matrix. The state transition model assumes that the estimated velocity is constant, and any changes are due process noise ($\boldsymbol{w_k}$). The noise vector $\boldsymbol{w_k}$ is modeled to be zero mean Gaussian white noise. In this study we consider two components contributing the elements of $\boldsymbol{w_k}$; i.e. pn_{ij}, the pixel range value process noise, and vn_{ij}, the axial range velocity. $\boldsymbol{Q_k}$ is the covariance matrix of the process noise vector. The second set of Kalman equations dictate how the current state is related to the current measurement made by the range camera. They are given by

$$\mathbf{z_k} = \mathbf{H_k} \mathbf{x_k} + \mathbf{v_k} \quad (6)$$

$$\mathbf{v_k} = \left[\ldots, \sigma^2_{(i-1),j}, \sigma^2_{i,j}, \sigma^2_{(i+1),j}, \ldots \right]^T \quad (7)$$

$$\mathbf{R_k} = \mathbf{v_k} \mathbf{v_k}^T \quad (8)$$

$$\mathbf{H_k} = \left[\mathrm{I} \mid \mathrm{O} \right] \quad (9)$$

where $\boldsymbol{z_k}$ is a vector representation of the measurement within the $m \times n$ window at time t_k, the vector $\mathbf{v_k}$ is the measurement error vector, and $\boldsymbol{R_k}$ is the measurement error covariance matrix. $\boldsymbol{H_k}$ is the state relationship matrix showing how the previously predicted value $\boldsymbol{x_k}$ is related to the measured value $\boldsymbol{z_k}$. Here, I is the $m \times n$ identity matrix and O is a zero matrix of the same size. The measurement vector contains only the pixel range values, since no velocity components can directly be measured. Figure 3 depicts the Kalman filter implementation details. The process and measurement noise vectors will remain constant through the Kalman prediction. The error covariance vector is continuously updated through the Kalman filtering process with an initial estimate of zero. The filter is given an error covariance matrix previous estimate $\mathbf{P}_k^-$ as the covariance of the current state.

If the velocity is determined to be either above or below a selected threshold, the filter estimate is considered to be invalid and iteration is terminated. The filter provides estimates of the error within the local $m \times n$ window. This covariance can be examined to help determine if the current window estimate

should be used. Furthermore, tracking or navigation can benefit from determining which part of the image is moving in free space by examining the covariance and the local velocity estimates.

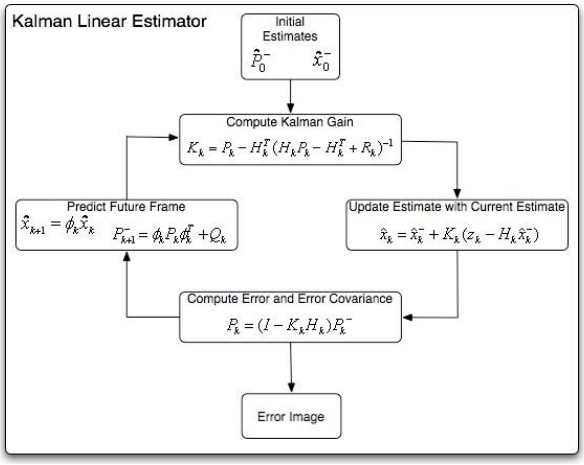

Fig. 3. Kalman linear estimator (from [16]).

3. EXPERIMENTAL RESULTS

The 3-D scene data used in this research is gathered using a SwissRanger SR 3000 camera [17] that senses the scene by gathering gray-scaled intensity data and *x*, *y*, *z* spatial coordinates in IR images (850nm) via time-of-flight measurements. These time-of-flight measurements provide an efficient way to collect 3-D scene data via phase-shift measurements of a reflected modulated signal [18]. Four projection images are created per view, three of which are for the spatial positions of scene points relative to the camera's image-plane coordinates. The fourth is used for the IR scene intensity map. The Swiss Ranger SR 3000 camera provides range measurements up to a distance of 7.5 meters with a frame resolution of 176x144 pixels and a maximum frame rate of 50 fps (frames per second). Since the image has a relatively high resolution, the Kalman filter equations are computed by using 3x3 non-overlapping subsets of the images in each frame independently. The predicted results are then combined to create a full 3-D scene image array. In our experiment the pixel noise (pn_{ij}) is assumed to be 0, and the velocity noise (vn_{ij}) is taken to be 1 m/s. The state transition matrix ϕ_k is adjusted for a 3x3 window based Kalman filter realization as an 18x18 matrix given by $\phi_k = \begin{bmatrix} I & I \\ 0 & I \end{bmatrix}$, where *I* is a 9x9 identity matrix and *0* is a 9x9 zero matrix. Figures 4 and 5 show the measured and predicted data for two consecutive frames 12 and 13, respectively, while Figures 6 and 7 depict the corresponding 3-D point cloud images. These figures illustrate the self-correcting nature of the Kalman filter. Although, the predicted image in Fig.8 is visibly distorted, the image in Fig.7 is much closer to the actual one.

The prediction error for the described method is calculated in the least mean square sense, which is given by

$$e = \sqrt{\frac{1}{M \cdot N} \sum \sum \left(\hat{f}_{k+1}(i,j) - f_{k+1}(i,j) \right)^2} \qquad (10)$$

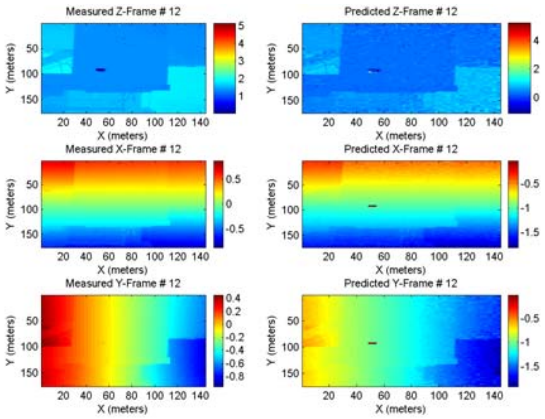

Fig. 4. x, y, z images for frame #12.

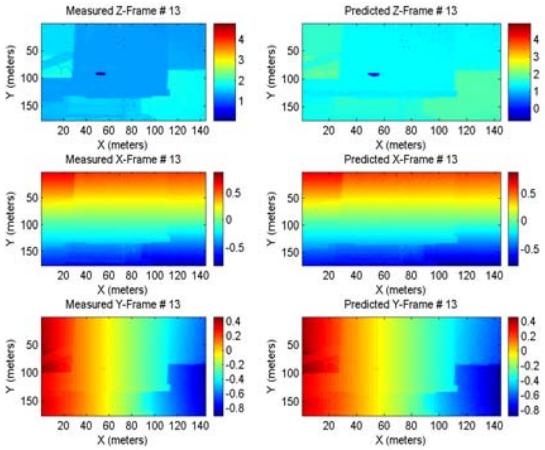

Fig. 5. x, y, z images for frame #13.

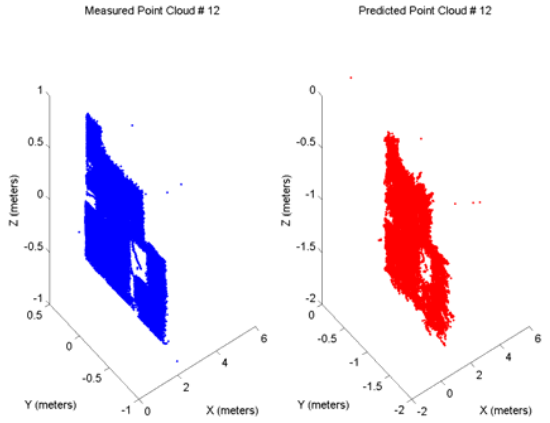

Fig. 6. Measured and predicted point clouds for frame #12.

where f denotes *x*, *y*, or *z* image values for the frame *k+1*. In Figure 8, the LMSE is depicted for a selected set of ten frames. It appears from this and additional tests that the initial value of the LMSE error can be as high as 4.2% and drops below 1% as prediction converges to the desired (actual) value. Once the algorithm begins to converge, the resulting error levels appear to become relatively flat. Furthermore, the results for all frames are drawn in Fig. 9 to give an overall assessment of the algorithm's performance. In all tests, once the filter adjusts itself, error, for all frame axes, falls to an acceptably low level. Remaining

fluctuations may be attributed to changes in the IR images and/or noise.

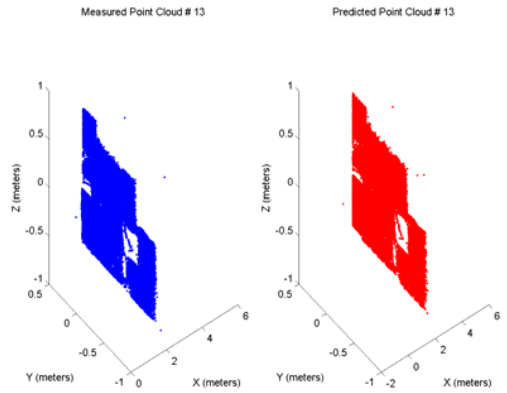

Fig. 7. Measured and predicted point clouds for frame #13.

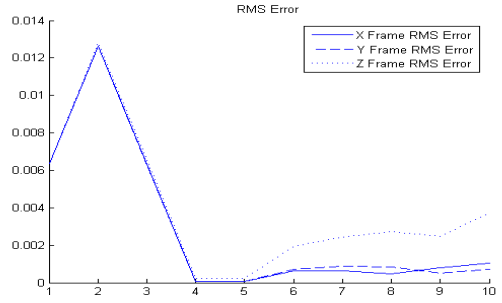

Fig. 8. LMSE error for frames 9-19.

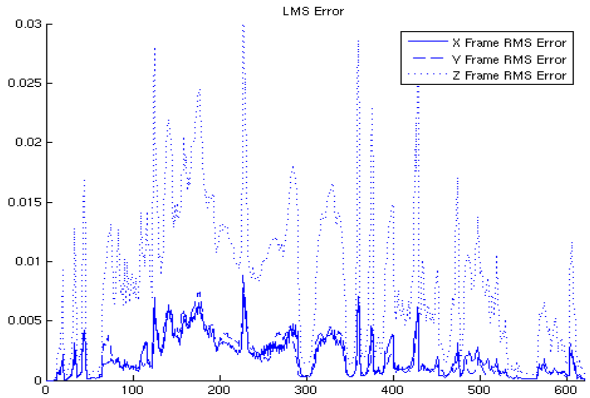

Fig. 9. LMSE measurement results for all frames used.

4. CONCLUSIONS

This work has the objective of predicting scenes as the camera or sensory device mounted on a mobile platform (e.g., walking robot, unmanned vehicle, low altitude aircraft, etc.) moves. Experimental results reported herein demonstrate that linear Kalman filtering based scene prediction can accurately estimate the next frames in range and intensity images to a certain degree of accuracy. The low LMSE error measurement, on the average of almost 1%, proves the reliability and robustness of this approach to IR data processing. The presented results are within the allowable error range of the low-cost cameras used for the experimentation. Future research is undertaken for non-linear camera motion and occlusion cases.

5. ACKNOWLEDGEMENTS

The authors acknowledge Jacob Campbell of AFRL for providing the Swiss Ranger data and Maarten Uijt de Haag of Ohio University for helping develop code for the algorithm.

6. REFERENCES

[1] J. Kim and J.W. Woods, "3D Kalman filter for image motion estimation," IEEET-IP, Vol.7, No.1, Jan.1998, pp.42-52.

[2] M. Irani and P. Anandan, "A unified approach to moving object detection in 2D and 3D scenes," IEEET-PAMI, Vol.20, No.6, June 1998, pp.577-589.

[3] E. B. Meier and F. Ade, "Object detection and tracking in range image sequences by separation of image features," Proc. IEEE Int. Con. on Intell. Veh., 1998, pp. 176-181.

[4] A. Hoover, et al., "Egomotion estimation of a range camera using the space envelope," IEEET-SMCB, Vol.33, No.4, Aug. 2003, pp.717-721.

[5] G. Vosselman and S. Dijkman, "3D building model reconstruction from point clouds and ground planes," Proc. Land Surface Map. & Char. Using Laser Altimetry, 2001, pp.37-43.

[6] R. Sagawa, et al., "Incremental mesh modeling and hierarchical object recognition using multiple range images," Proc. IROS 2000, Vol.1, Oct. 31-Nov. 5, 2000, pp.88-95.

[7] B.L. Rozovskii and A. Petrov, "Optimal nonlinear filtering for track-before-detect in IR image sequences," Proc. SPIE Vol.3809, 1999.

[8] B.J. King, et al., "Registration of multiple range scans as a location recognition problem: Hypothesis generation, refinement and verification," Proc. Fifth Int. Conf. 3DIM, June 13-16, 2005, pp.180-187.

[9] H.S. Sawhney, et al., "Independent motion detection in 3D scenes," IEEET-PAMI, Vol.22, No.10, 2000, pp.1191-1199.

[10] K. J. Jager, et al., "Automatic 3D object pose estimation in IR image sequences for forward motion applications," Proc. SPIE Vol.5426, pp.37-45, 2004.

[11] S.I. Roumeliotis and G.A. Bekey, "SEGMENTS: A layered, dual-Kalman filter algorithm for indoor feature extraction," Proc. Int. Conf. Intel. Robots & Syst., 2000, pp. 454-461.

[12] S. Kim and I. S. Kweon, "3D target recognition using cooperative feature map binding under Markov chain Monte Carlo," Pat. Rec. Let., Vol.27, Iss.7, pp.811-821, May 2006.

[13] C. Ibarra-Castanedo, et al., "IR image processing and data analysis," IR Phy.&Tech., Vol.46, Iss.1-2, pp.75-83, 2004.

[14] T. Tsao and Z. Wen, "Image-based target tracking through rapid sensor orientation change," OE, Vol.41, pp.697-703, 2001.

[15] M.A. Zaveri, et al., "Tracking multiple maneuvering point targets using multiple filter bank in infrared image sequence," Proc. ICASSP '03, Vol.2, pp.409-412, 2003.

[16] R.G. Brown and P.Y.C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, New York, 1992.

[17] Swiss Ranger SR-300, www.swissranger.ch/main.php

[18] T. Oggier, et al., "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger^{TM})," SPIE Proc Vol. 5249-65, 2003.

[19] K. F. Manizade, et al., "Stereo cloud heights from multispectral IR imagery via ROI segmentations," IEEET-GRS, Vol.44, No.9, pp.2481-2492, 2006.