

# A Filter Method for Pose Estimation of Maneuvering Target

Feili Hou<sup>1,2</sup>, Feng Zhu<sup>1</sup>

<sup>1</sup> Robotics Laboratory  
Shenyang Institute of Automation, Chinese Academy of Sciences  
Shenyang, P.R.China

<sup>2</sup> Graduate School of the Chinese Academy of Sciences  
Beijing, P.R.China

**Abstract**—A filter method is presented for 3D pose (position and orientation) estimation of an arbitrary moving target from a monocular image sequence. Problem considered here is error propagation from image features to pose parameters. To this end, we have derived a combination of two filter schemes. The first filter scheme uses maneuver detection technique, in which the criterion of an optimal detector is deduced, two detectors suitable for fast and slow maneuvers are analyzed, and limited memory filtering is adopted for maneuver correction. In the second filter, a robust model is established via Lagrange interpolation and numerical integration, and the possibility of filter divergence is avoided by employing an adaptively estimated fading factor. Finally, generalized pseudo Bayes algorithm is employed to combine the two filter models for better filter performance. Superior to previous approaches that were limited to slow and smooth motion, this filter method is applicable for a maneuvering target that acts in an unknown manner. Moreover, it is suitable for real-time environment in terms of the following aspects: the structure of the filter is simplified by avoiding the need of EKF, the computational cost is much reduced by running six filters on the six pose sequences in parallel; the process to extract the feature points is simplified due to the predicted feature locations; time efficiency is increased because the sequence of image frames is allowed to be at unequal intervals. Simulation results and implement of a system for multi-mobile robots formation show the capacity of this algorithm.

**Keywords**—pose estimation; kalman filter; maneuver detection; numerical analysis; filter fusion; robots formation

## I. INTRODUCTION

Recently, a significant amount of work has been done to the determination of the 3D pose of a moving target from a monocular sequence of images. This is because that pose estimation can offer many important applications in a wide range of areas, including mobile robot navigation, target tracking, and motion prediction etc. One of the central problems in the 3D pose estimation noted by several researchers [1,2] is that: perturbation of all kinds of

noises embedded in the images results in errors in 3D reconstruction, and thereby, causes serious degradation in pose estimation. Based on the fact that a long image sequence can provide more information to suppress noise, application of the extended kalman filter (EKF) has been widely used to improve the numerical accuracy of the pose parameters. However, there are still some limitations on the ability of these methods [5-10]. Generally, the modeled motion was assumed to have constant acceleration/ velocity in the translational and rotational components, which made the filter only suitable to application where the object motion is slow and smooth. In other applications where the object may travel through a sudden movement or a step discontinuity, more work is needed.

To increase the robustness of the overall method, a new filter approach is presented here for accurate 3D pose estimation of a maneuvering target from a monocular image sequence, with no need of prior knowledge about its trajectory. The overall filter method is implemented by combining two filter modules: the maneuver detection based filter (MDF) and the numerical analysis based filter (NAF). In MDF, the criterion of an optimal detector has been deduced, and two detectors suitable for fast and slow maneuvers are analyzed respectively. Once maneuver is declared, limited memory filtering is introduced for maneuver correction. Superior to the correction method of acceleration input estimation [13,14], better filter performance is achieved because the opposite influence caused by poor estimation of the starting time of maneuver is avoided. In NAF, a robust model applicable to diverse dynamics is established by Lagrange interpolation and numerical integration. And, a fading factor is imported in the covariance matrix of the prediction error to overcome the phenomenon of filter divergence caused by truncation errors. This factor can be estimated adaptively. Finally, the two filter modules are combined by the generalized pseudo Bayes algorithm. Combination of the two filter modules has greatly increased the accuracy of the filter, especially at the maneuvering time when the target goes through a sudden change.

As well as reducing the effect of noise, this filter has strong ability for real time application in terms of following aspects. First, establishment of linear measurement equation avoids the need of EKF, which makes the structure of the filter much simpler and computationally

---

This work is supported by the National High Technology Research and Development Program of China (NO. 2002AA401001-4A)

more efficient. Second, running six separate filters for the six pose sequences in parallel helps reduce the computation time. Third, the sampling intervals between the image frames are allowed to be unequal, which would increase the efficiency of the 3D motion estimation system. Last, based on the predicted pose parameters provided by the filter, image plane regions to which feature extraction is restricted can be acquired. This allows the vision sensor system to process only small window areas in the image plane to obtain the required feature points, and thus, brings a significant reduction in the image processing time.

In the following text, the process of 3D pose estimation from monocular vision is presented in section II. In section III, linear measurement equation is derived. Section IV describes the first filter scheme using maneuver detection technique, while section V describes the second filter scheme based on the estimate model of Lagrange interpolation and numerical integration. After the combination of the two filter modules is given in section VI, simulation results and practical application of a system for multi-mobile robots formation are presented in section VII. Finally, a conclusion is made in section VIII.

## II. POSE ESTIMATION

This section will describe the reconstruction of 3D object pose by the model-based monocular vision method. Geometry of a monocular vision system for pose estimation is illustrated in Fig. 1. A pinhole camera model is used for mapping, and no less than three features points are selected for pose estimation. The camera is assumed to be stationary in inertial coordinates, and the imaging model is that of central projection. An arbitrary body coordinate frame is attached to the object to define the orientation and position of the object. For the pose estimation, following process should be accomplished: the extraction of feature points from 2D images, the matching of corresponding feature points between 2D image planes and 3D space, and the transformation between the two frames.

At each measurement instant, the relative pose of the object with respect to the camera is estimated by following method: the transformation of the  $j^{\text{th}}$  feature measurement in the object frame,  $P_j^o = [x_j^o, y_j^o, z_j^o]^T$ , to the camera frame,  $P_j^c = [x_j^c, y_j^c, z_j^c]^T$ , can be performed by a translation  $T = [x, y, z]^T$  and a rotation  $R(\alpha, \beta, \gamma)$ :

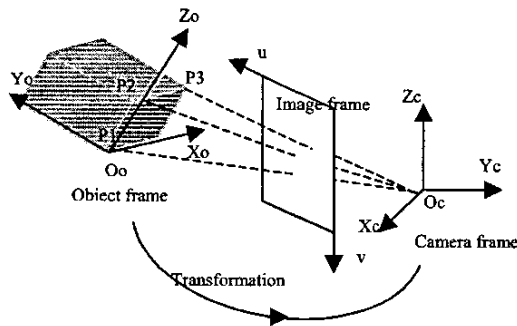


Figure 1. Geometry of a monocular vision system

$$P_j^c = T + RP_j^o \quad (1)$$

From the point projection camera model, the corresponding projection feature point locations in the image frame,  $[u_j, v_j]^T$ , are obtained by

$$u_j = K_u x_j^c / z_j^c, v_j = K_v y_j^c / z_j^c \quad (2)$$

where  $K_u$  and  $K_v$  are the calibrated camera interior parameters. Since the spatial locations of the features points is known beforehand, we may find value of  $P_j^c$  according to the model-based monocular vision method. So, the rotation matrix  $R$  and the translation vector  $T$  in (1) can be acquired by the least square method. Namely, the pose parameters, including the orientation parameters (roll  $\alpha$ , pitch  $\beta$ , yaw  $\gamma$  angle) and position parameters (forward  $x$ , lateral  $y$ , and vertical  $z$  translation), are obtained.

## III. MEASUREMENT MODEL

In conventional measurement model, noisy image feature vector was selected as measurements. However, considering its nonlinear relationship with the pose parameters, here we choose the estimated pose parameters, instead of the image feature vector, as a measurement. The discrete measurement equation relating nominal pose parameters  $W' = [x', y', z', \alpha', \beta', \gamma']^T$  and the true ones  $W = [x, y, z, \alpha, \beta, \gamma]^T$  at the  $k^{\text{th}}$  sampling step is

$$W_k' = W_k + V_k \quad (3)$$

where  $V = [n_x, n_y, n_z, n_\alpha, n_\beta, n_\gamma]$  is the noise component.

Because the images taken from a longer distance have larger noise-to-signal ratios in 2D image frames, they will generate larger errors in terms of object position and orientation in 3D space. It is therefore more realistic to assume that the measurement noise in 2D image-plane points, rather than the pose noise of  $V_k$ , is governed by the same statistics for all the time. Any errors incurred during the extraction of feature points are incorporated into the 2D image noise as independent white-Gaussian noise with the same statistics for all the frames. From the mathematical relation between the image features to pose parameters, we can get the effect of 2D noise on the object orientation and position in 3D space, and derive the maximum error for each pose parameters. Assume that measurement noise  $V_k$  is zero mean Gaussian process. According to the 3-sigma rule, if the maximum value of the  $i^{\text{th}}$  element in  $V_k$  at the  $k^{\text{th}}$  sampling step is  $M_k^i (i=1..6)$ , its covariance can be approximated as  $(M_k^i/3)^2$ .

In the vision based 3D motion estimation system, if the intervals between successive frames are predetermined to be equal, much time will be wasted on waiting until the predetermined sampling time has come. On the contrary, if

the camera can work immediately once various processes have been completed, better efficiency will be attainable. Therefore, to increase the ability of real time application of the whole system, the dynamic models established in the following two filters modules are designed to be fit for changing sampling period.

As well as the measurement equation, the dynamic equations to be constructed next are linear both in MDF and NAF. This allows for an accurate estimation of pose parameters through linear Kalman filtering, and thus, has drastically reduced the computational complexity. Moreover, we can run six filters on the six pose sequences in parallel to reduce the computational time.

Next, we will give two filter modules of MDF and NAF respectively, and further, present the filter fusion process. Filtering on the forward translation  $x$  is chosen as representative. Other pose parameters are filtered in the similar way and therefore omitted.

#### IV. MANEUVER DETECTION BASED FILTER

##### A. Models

The dynamic equation is given by a three-state model

$$X_{k+1} = A_k X_k + W_k^x \quad (4)$$

where state vector  $X_k = [x(t_k) \quad x'(t_k) \quad x''(t_k)]^T$ ,

transition matrix  $A_k = \begin{bmatrix} 1 & t_{k+1} - t_k & (t_{k+1} - t_k)^2 / 2 \\ 0 & 1 & t_{k+1} - t_k \\ 0 & 0 & 1 \end{bmatrix}$ , and

dynamic noise  $W_k^x$  is assumed to be zero mean Gaussian noise with covariance  $Q_k^x$ . When a maneuver occurs, the target process becomes

$$X_{k+1} = A_k X_k + W_k^x + B_k U_k \quad (5)$$

at the precise instant,  $k=n$ , at which the maneuver occurs. Here the input acceleration is  $U_k$ , and the input matrix  $B_k = [(t_{k+1} - t_k)^2 / 2 \quad t_{k+1} - t_k \quad 1]^T$

From (3), the discrete measurement equation on  $x$  can be expressed by

$$Z_k = C X_k + V_k^x \quad (6)$$

where  $Z_k$  is the noisy estimation of  $x(t_k)$ ,  $C = [1 \quad 0 \quad 0]$ ,  $V_k^x$  is measurement noise, assumed to be zero mean Gaussian noise with time-variant covariance  $R_k^x$ .

##### B. Maneuver Detection

###### 1) Criterion for an optimal detector

Our goal is to detect the existed maneuver as soon as possible. So, an optimal detector with the maximum

probability of maneuver detection is explored next.

Suppose that the maneuver starts at the sampling step of  $n$ , with acceleration input of  $U_n$ . We let  $\tilde{X}_{k|k-1}$  denote the predicted estimate assuming no maneuver, while  $\hat{X}_{k|k-1}$  denote the prediction when  $U_n$  is already known. It can be derived that the relation between the biased innovation,  $r_k = Z_k - C\tilde{X}_{k|k-1}$ , and the unbiased one,  $e_k = Z_k - C\hat{X}_{k|k-1}$ , is

$$r_k = M_k^n U_n + e_k \quad (7)$$

where  $M_k^n = \begin{cases} 0 & k \leq n \\ CB & k = n+1 \\ C[\prod_{l=n+1}^{k-1} A_l (I - K_l C)] B_n & k > n+1 \end{cases}$ . If the

distribution of  $e_k$  is normally distributed as a zero mean normal random variable with variance  $\sigma_k^2$ , we can get that  $r_k$  is normally distributed as

$$N(M_k^n U_n, \sigma_k^2) \quad (8)$$

Let  $d$  as the number of measurements used to estimate the detector, then the maneuver detector composed of standardized weighted sum of innovation has the form of

$$H_k = \frac{\sum_{i=1}^d h_i r_{k-d+i}}{\sqrt{\sum_{i=1}^d h_i^2 \sigma_{k-d+i}^2}} \quad (9)$$

, in which  $h_i$  is the weight coefficient. If  $H_k$  exceeds a certain threshold, we accept the hypothesis that a maneuver has taken place. From (8) and (9), we get the distribution of  $H_k$  as:

$$N\left(\frac{\sum_{i=1}^d h_i M_{k-d+i}^n U_n}{\sqrt{\sum_{i=1}^d h_i^2 \sigma_{k-d+i}^2}}, 1\right) \quad (10)$$

Define

$$J_k = \frac{\sum_{i=1}^d h_i M_{k-d+i}^n U_n}{\sqrt{\sum_{i=1}^d h_i^2 \sigma_{k-d+i}^2}} \quad (11)$$

It can be proven that under a given probability of false alarm, the probability of detection will be maximum if  $|J_k|$  is maximal. Therefore, by differentiating  $J_k$ , we reach the conclusion that under the condition of

$$h_i^{opt} = M_{k-d+i}^n \sigma_{k-d+i}^{-2} / S \quad (12)$$

(S is a random constant),  $H_k$  in (9) is the optimal detector.

## 2) Detectors for quick and slow maneuvers

In fact, the starting step of maneuver is usually not obtainable, so  $M_k^p$  cannot be accurately determined. Hereby, we need to probe sub-optimal detectors for different circumstances.

Analyzing the input estimation based detector in [13,14]

$$H_k = \hat{U}_k / \sqrt{\text{Var}(\hat{U}_k)} \quad (13)$$

where  $\hat{U}_k$  is the estimated input at the sampling step of k

$$\hat{U}_k = M_{k-d+1}^{k-d} \sigma_{k-d+1}^{-2} r_{k-d+1} / \sum_{j=1}^d (M_{k-d+j}^{k-d})^2 \sigma_{k-d+j}^{-2} \quad (14)$$

, we find that the shorter the maneuver detection delay (i.e. the distance between k-d and n), the closer the detector of (13) is to the optimal detector. So, the detector proposed in [13,14] is suitable for fast maneuver.

In slow maneuver, since the detection delay is long, we can deem that  $M_{k-d+i}^p$  ( $i=1\dots d$ ) equals to each other. And  $\sigma_{k-d+i}^{-2}$  can also be approximated as the same. Hereby, it is reasonable to approximate the weight coefficient in (12) as  $h_i^{opt} \approx 1$ . So, the detector

$$H_k = \sum_{i=1}^d r_{k-d+i} / \sqrt{\sum_{i=1}^d \sigma_{k-d+i}^2} \quad (15)$$

is suitable for slow maneuver.

## C. Filer Update

Once maneuver is declared, limited memory filtering is adopted here for maneuver correction. Compared with the acceleration input estimation method [13,14], this correction scheme is superior because it does not need to estimate the starting time of maneuver. Therefore, adverse influence caused by its poor estimation can be avoided. The update process is: We choose N as the memory length. Once a maneuver is detected at the sampling step of k, the estimated state  $\hat{X}_k$  and its error covariance matrix  $\hat{P}_k$  will be corrected using the observations between the sampling interval of  $k-N$  and  $k$ ,  $\theta_N = \{Z_{k-N+1} \dots Z_k\}$ , and replaced by  $\hat{X}_{k|\theta_N}$  and  $\hat{P}_{k|\theta_N}$ :

$$\begin{cases} \hat{X}_{k|\theta_N} = (\hat{P}_k^{-1} - \hat{P}_{k|k-N}^{-1})^{-1} (\hat{P}_k^{-1} \hat{X}_k - \hat{P}_{k|k-N}^{-1} \hat{X}_{k|k-N}) \\ \hat{P}_{k|\theta_N} = (\hat{P}_k^{-1} - \hat{P}_{k|k-N}^{-1})^{-1} \end{cases} \quad (16)$$

## V. NUMERICAL ANALYSIS BASED FILTER

### A. Lagrange Interpolation and Numerical Integration

From the Newton Law of motion, the macroscopical motion of  $f(t)$  satisfies two characters as below: (1) There are numerable isolated  $n^{\text{th}}$ -order jump-points ( $n > 1$ ). (2) The  $n^{\text{th}}$ -order derivative of  $f(t)$  between two ordinal jump-points exists and all derivations are bounded. Consequently, it is reasonable to use Lagrange interpolation and numerical integration to describe the object 3D motion.

Lagrange 2-degree Interpolation with unequal intervals is considered here. The interpolating polynomial passing through the three points  $(t_{k-2}, f(t_{k-2}))$   $(t_{k-1}, f(t_{k-1}))$  and  $(t_k, f(t_k))$  is given by

$$g(t) = f(t_{k-2})L_{k-2}(t) + f(t_{k-1})L_{k-1}(t) + f(t_k)L_k(t) \quad (17)$$

where the quotients are

$$\begin{cases} L_{k-2}(t) = (t-t_{k-1})(t-t_k) / [(t_{k-2}-t_{k-1})(t_{k-2}-t_k)] \\ L_{k-1}(t) = (t-t_{k-2})(t-t_k) / [(t_{k-1}-t_{k-2})(t_{k-1}-t_k)] \\ L_k(t) = (t-t_{k-2})(t-t_{k-1}) / [(t_k-t_{k-2})(t_k-t_{k-1})] \end{cases}$$

If  $t_{k-2}, t_{k-1}, t_k$  are distinct numbers in the interval  $[t_a, t_b]$ , and  $f \in C^3[t_a, t_b]$ , the truncation error between  $f(t)$  and  $g(t)$  is:

$$\omega(t) = \frac{f^{(3)}(\varepsilon)}{3!} (t-t_{k-2})(t-t_{k-1})(t-t_k), \quad \varepsilon \in [t_a, t_b] \quad (18)$$

Based on the interpolating polynomials given in (17), the integral of  $f(x)$  over  $[t_c, t_d]$  ( $t_a \leq t_c < t_d \leq t_b$ ) can be approximated by the method of numerical quadrature:

$$\begin{aligned} \int_c^d f(t) dt &= f(t_{k-2}) \int_c^d L_{k-2}(t) dt + f(t_{k-1}) \int_c^d L_{k-1}(t) dt \\ &+ f(t_k) \int_c^d L_k(t) dt + \int_c^d \omega(t) dt \end{aligned} \quad (19)$$

### B. Models

Replacing the variables of  $f(t)$  and  $t_c$  in (17-19) with  $x'(t)$  and  $t_k$  respectively, we can get the following expression:

$$\begin{aligned} x'(t) &= x'(t_{k-2})L_{k-2}(t) + x'(t_{k-1})L_{k-1}(t) + x'(t_k)L_k(t) + \omega(t) \\ x(t) &= x(t_k) + \int_k^t x'(h) dh \\ &= x(t_k) + x'(t_{k-2}) \int_k^t L_{k-2}(h) dh + x'(t_{k-1}) \int_k^t L_{k-1}(h) dh \\ &+ x'(t_k) \int_k^t L_k(h) dh + \int_k^t \omega(h) dh \end{aligned} \quad (t > t_k) \quad (20)$$

Let  $X(k) = [x(t_k), x'(t_k), x'(t_{k-1}), x'(t_{k-2})]^T$  be the system state vector, the discrete dynamic model is accordingly derived as

$$X_{k+1} = A_k X_k + W1_k^x + W2_k^x \quad (21)$$

in which

$$A_k = \begin{bmatrix} 1 & \int_k^{k+1} L_k(t) dt & \int_k^{k+1} L_{k-1}(t) dt & \int_k^{k+1} L_{k-2}(t) dt \\ 0 & L_k(t_{k+1}) & L_{k-1}(t_{k+1}) & L_{k-2}(t_{k+1}) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ is the}$$

transition matrix,  $W1_k^x = \left[ \int_k^{k+1} \omega(t) dt \quad \omega(t_{k+1}) \quad 0 \quad 0 \right]^T$  is the truncation error, and  $W2_k^x$  is the system noise.

The expression precision of (21) is impaired by the truncation error of  $W1_k^x$ , so its effect should be evaluated. According to the kinematics mechanism of macroscopical motion, it is known that between two ordinal jump-points, higher-order derivatives of  $x(t)$  are bounded in finite value. Moreover, the sampling intervals here can be controlled to be tiny enough. Therefore, it is reasonable to consider the truncation error as modeling error with bounded energy. Disturbance of bounded energy brings bounded estimate error for convergent estimate algorithms (BIBO system).

Many structured dynamic models are confined to some compatible dynamic modalities. Once modality varies, incompatible between model and process will occur. Whereas, the model of (21) can describe motion of diverse modality and avoid such incompatible. Therefore, it is a "flexible" expression tool to describe unknown function without its analytic formula. Because the estimate model of (21) connotes some physical background, it usually satisfies the condition of uniformly controllable and observable. Therefore, it is easy to construct convergent estimate algorithm [3].

As for the measurement equation, it can be expressed by

$$Z_k = CX_k + V_k^x \quad (22)$$

where  $C = [1 \ 0 \ 0 \ 0]$ ,  $Z_k$  and  $V_k^x$  are the same as the value in (6).

### C. Filter Divergence Avoidance

Existence of the truncation error makes it is difficult to accurately determine the dynamic noise covariance  $Q_k^x$  in (21). Poor estimation of  $Q_k^x$  will degrade the accuracy of the filter performance, and may even cause divergence. To avoid the possibility of divergence, we import a fading factor  $S_k^x$  here. The divergence avoidance scheme is:

Let  $\delta_k = Z_k - C\hat{X}_{k|k-1}$  be the innovation, with

covariance  $E[\delta_k^T \delta_k] = CP_{k|k-1}C^T + R_k^x$ , and  $\gamma$  be the scale factor, if

$$\delta_k^T \delta_k > \gamma [CP_{k|k-1}C^T + R_k^x] \quad (23)$$

we suppose that the filter is divergent. Then, the covariance matrix of the prediction error

$$P_{k|k-1} = A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1}^x \quad (24)$$

will be corrected by

$$P_{k|k-1} = S_k^x A_{k-1}P_{k-1}A_{k-1}^T + Q_{k-1}^x \quad (25)$$

The fading factor  $S_k^x$  acts by reducing the weight of the old observations and increasing that of the new observations. It can be determined automatically according to the innovations:

$$S_k^x = \frac{1}{m} \text{tr} \{ [\delta_k \delta_k^T - CQ_{k-1}^x C^T - R_k] [CA_{k-1}P_{k-1}A_{k-1}^T C^T]^{-1} \} \quad (26)$$

, in which  $m$  is the observation dimension.

## VI. FILTER FUSION

In order to get higher tracking quality, generalized pseudo-Bayes algorithm is introduced here for combination of the two filter modules. This fusion approach is composed of the two above filter modules, a model probability evaluator, and an estimate combiner at the output of the filters.

Let the notation  $M_k^i$  stands for "model  $i$  ( $i=1,2$ ) in effect during the period ending at the sampling step of  $k$ ", which is assumed to be a finite Markov chain with the transition probability matrix of  $H$ . And, define  $N^k = \{Z_1 \dots Z_k\}$  as "the cumulative data at the sampling step of  $k$ ",  $\mu_k^i = P(M_k^i | N^k)$  as "the model probability of the  $i^{\text{th}}$  filter at sampling step of  $k$ ". Then, one cycle of the fusion approach consists of the following:

- Filter iteration:

Each of the pairs  $X_{k-1}^i$  and  $P_{k-1}^i$  are used as input to the filter matched to  $M_k^i$ . Time-extrapolation yields  $X_{k|k-1}^i$  and  $P_{k|k-1}^i$ , and measurement update provides  $X_k^i, P_k^i$ .

- Likelihood function update:

The likelihood function corresponding to filter  $i$  is computed according to

$$\Delta_k^i = P(Z_k | M_k^i N^{k-1}) = \frac{1}{\sqrt{2\pi}} |\Sigma_k^i|^{-1/2} \exp\left(-\frac{1}{2} r_k^{i^T} \Sigma_k^{-1} r_k^i\right) \quad (27)$$

where  $r_k^i$  is innovation of the  $i^{\text{th}}$  filter at sampling step of  $k$ ,

whose covariance is  $\Sigma_k^i$ .

- Model probability update:

From the Chapman-Kolmogorov equation for the Markov chain, we easily obtain

$$P(M_k^i | N^{k-1}) = \sum_m H_{im} \mu_{k-1}^m \quad (28)$$

Then, the model probability of each filter is updated according to the Bayes theorem

$$\mu_k^i = \frac{P(Z_k | M_k^i N^{k-1}) P(M_k^i | N^{k-1})}{\sum_j P(z_k | M_k^j N^{k-1}) P(M_k^j | N^{k-1})} = \frac{\Delta_k^i \sum_m H_{im} \mu_{k-1}^m}{\sum_j \left( \Delta_k^j \sum_m H_{jm} \mu_{k-1}^m \right)} \quad (29)$$

- Output Fusion:

On the basis of the law of total probability, Output is expressed as

$$X_k = \sum_i \mu_k^i X_k^i \quad (30)$$

## VII. RESULTS

### A. Simulation Results

A computer simulation was performed to test the performance of our filter approach. In the computer simulation, image generation, feature extraction, and 3D object motion were implemented along with the Kalman filter algorithm. Simulation data were obtained from the scenario shown in Fig. 2. In order to verify the validity of this approach for general conditions, pose parameters were performed with different motion velocities and trajectories. The time intervals between two consecutive measurements were generated randomly by the computer between 0.1s ~ 0.15s. At every sampling time, the noisy pose measurements were made in the following way: noise free 3D feature points generated by a computer program were projected onto the image plane. Then, white independent Gaussian noise with predetermined distribution was added to the true 2D image coordinate. With the 2D noisy features, a reconstruction of 3D pose was carried out by the model-based monocular vision method to simulate noisy measurements of pose parameters. A Monte Carlo

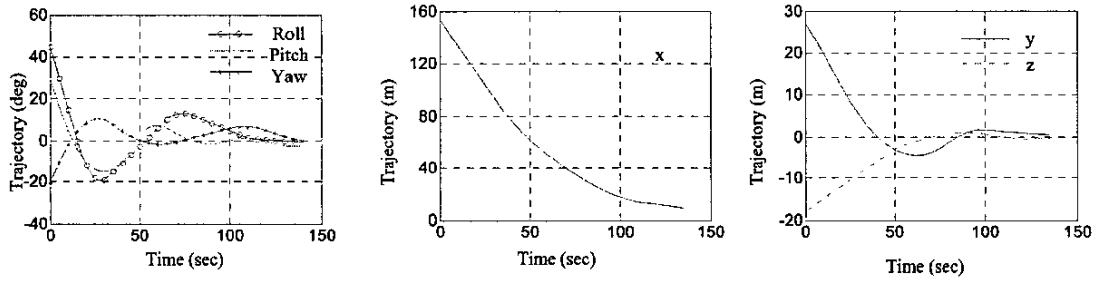
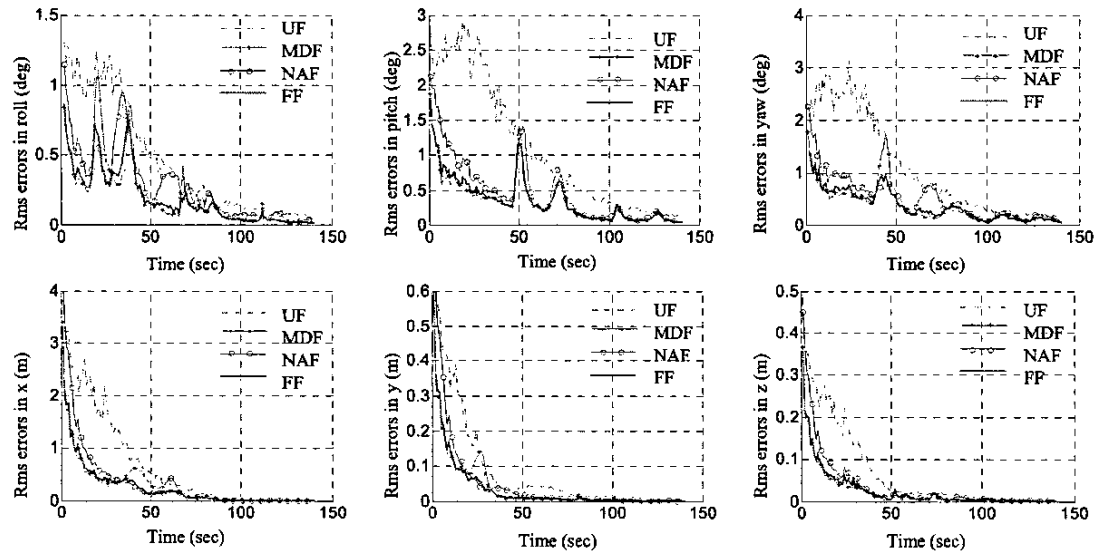


Figure 2. Object 3D motion trajectories in simulation



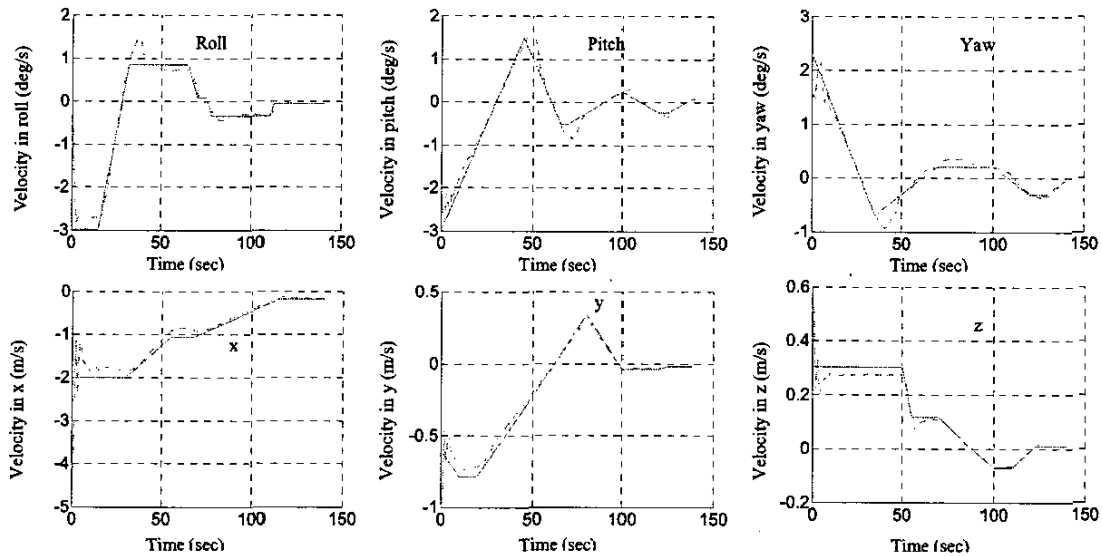


Figure 4. Estimates in rotational and translational velocities

simulation of 50 runs was obtained for the filter algorithm, with the same noise covariance for the image coordinates. Simulation results were given in following figures.

Fig. 3 shows the rms errors in both rotation and translation parameters under the maneuver detection based filter (MDF), the numerical differentiation based filter (NAF) and the fusion of the two filters (FF). The dotted curve (UF) denotes the rms errors before filtering. It is observed from the simulation that the errors in 3D space are proportional to the forward translation  $x$  (i.e., the distance between the object and the camera). The reason is that, as the object moves close to the camera, the size of the image expands, then the ration of the 2D measurement noise level to the image size of the object decreases, and consequently, less errors in 3D space are induced. Fig. 3 shows that both of the two filter schemes have decreased the rms errors. After combining the two filters, the peaks of rms errors are further decreased at the maneuvering time. Therefore, it has been verified that after filter fusion, better pose estimation will be achieved, especially during the time that the target motion goes through a sudden change.

The performances of our Kalman filter on rotational and translational velocities estimation are illustrated by Fig. 4. The solid lines represent the actual velocities of the motion shown in Fig. 2, and the dotted lines represent the velocity estimated by the Kalman filter. We can see from these figures that both rotational and translational velocities can be estimated very closely to the true values. Even when the object travels through a sudden change in its 3D motion, the filter experiences corresponsive transition and then recovers tracking in a short period.

### B. Experimental Implementation

Real-time experiment on vision-based moving in formation by three mobile robots was performed to demonstrate the validity and performance of our filter

approach. The experimental setup is shown in Fig. 5. Two sets of feature points were setup on Robot I and Robot II, while two cameras were mounted on Robot II and Robot III respectively. The three robots move in a triangular formation. Robot I served as a leader and goes first to provide moving plans to Robot II, and Robot III tracked the motion of Robot II. In this experiment, only three pose parameters of yaw angle, forward and lateral translation, were necessary to be processed.

Robot II tracked the motion of Robot I by detecting its relative pose with respect to Robot I. One cycle of the tracking process consisted of the following: After completion of the last tracking cycle, the camera that was attached on Robot II acquired the picture of Robot I. From the time length between the current image sampling time and the last one, we can determine the transition matrix for the dynamic equation. The regions to which feature extraction was restricted can also be computed based on the predictive ability of Kalman filter. These new window

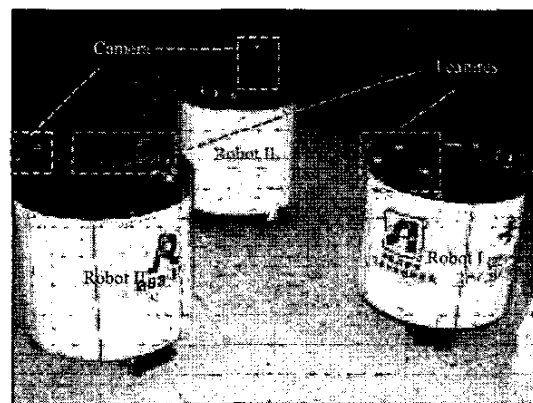


Figure 5. Setup of multi-robots formation

areas were used by the image processor to extract feature points for their image location measurements. Then, after feature extraction, we could derive independent relative pose measurements between Robot I and Robot II by the model-based monocular vision method. Noisy pose measurements were input to the filter to produce much more confident 3D motion estimates, including pose parameters and their velocities. Output signals of the filter were fed back to the controller of Robot II for its motion planning. So, Robot II would take an action such as proceeding forward/back or turning left/right to keep the formation. The above procedure was repeatedly continued. Robot III worked in the same way to track Robot II. Thus, the formation was kept all along.

It was observed that, thanks to the more accurate pose estimates provided by the filter, the tracking robots repositioned more quickly and accurately. Moreover, with the predicted pose parameters provided by the filter, feature point locations can also be predicted. Hereby, the process to extract the feature points became much simplified, and the image processing time was greatly reduced. Besides, because time intervals between two image sampling steps are allowed to be unequal, the cameras might catch the picture immediately after other processes have been completed, and no time was wasted on waiting for the predetermined periodic sampling time. As a result, the robot system's time efficiency was increased.

#### VIII. CONCLUSION

In this paper, a robust filter algorithm for more accurate estimation of the 3D relative pose of a random moving object from a monocular image sequence has been proposed. Linear measurement equations are derived, two filter algorithms based respectively on the maneuver detection method and the numerical analysis technique are improved, and better filter performance is acquired by combining the two filter modules. This filter method provides robust 3D pose estimation for a maneuvering object without prior knowledge about its trajectory. Moreover, it is very suitable for real-time environment in terms of many aspects: the computational cost is much reduced by running six filters on the six sequences in parallel; the image processing time is reduced by the predictive feature locations; no time is wasted on waiting because the sampling intervals is not need to be the same. Both computer simulation and real-time experiment results show that this filter will lead to a more robust and autonomous motion-tracking scheme.

#### REFERENCES

- [1] J.K. Aggarwal, "Motion and time-varying imagery-An overview," Proc. of IEEE Workshop on Motion: Representation and Analysis, Kiawah Island, South Carolina, May 1986.
- [2] T.S. Huang et. Al. "Motion detection and estimation from image sequences: Some preliminary experimental results," Proc. of IEEE Workshop on Motion: Representation and Analysis, Kiawah Island, South Carolina, May 1986.
- [3] P.Z. Jia, Z. T. Zhu, Optimal estimation theory and application, CN: The Publishing Company of Science, Beijing, 1984.
- [4] Feili Hou, Feng Zhu, "Limited memory filter for 3D relative position and orientation estimation of maneuvering target," Proceeding of the 5<sup>th</sup> World Congress on Intelligent Control and Automation, in press.
- [5] S. Lee, and Y. Kay, "A Kalman filter approach for accurate 3-D motion estimation from a sequence of stereo images," CVGIP, vol. 54, no. 2, Sep. 1991, pp. 244-258.
- [6] G.S. Young, and R. Cheelappa, "3D motion estimation using a sequence of noisy stereo images," IEEE Trans., vol. 12, no. 8, Aug. 1990, pp. 735-59.
- [7] J. Wang, and William J. Wilson. "3D relative position and orientation estimation using kalman filter for robot control," IEEE International Conference on Robotics and Automation, Nice, France, May 1992, pp.638-2645.
- [8] TJ Broida, S. Chandrashekhar, and R. Chellappa, "Recursive 3-D motion estimation from a monocular image sequence," IEEE Transactions on Aerospace and Electronic Systems, vol. 26, no. 4, pp. 639-656, July 1990.
- [9] TJ Broida, and R. Chellappa, "Estimation of object motion parameters from noisy images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 1, pp. 90-99, Jan. 1986.
- [10] M. Ficocelli and F. Janabi-Sharifi, "Adaptive filtering for pose estimation in visual servoing," Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, pp. 19-24, November 2001.
- [11] H.R. Zhou, Z.L. Jing,, P.D. Wang. Tracking of maneuvering targets. CN: National Defense Industry Press, Beijing, 1991.
- [12] C.Q. Xiang, T. Tian. Introduction of discrete estimation. CN: Harbin Ship&Boat Engineering Institute Press, Beijing, 1989.
- [13] Y.T. Chan, A.G. Hu, J.B. Plant. "A kalman filter based tracking scheme with input estimation," IEEE Transactions on Aerospace and Electronic Systems, vol. 15, no. 2, Mar. 1979, pp. 237-244.
- [14] P.L. Bogler, "Tracking a maneuvering target using input estimation," IEEE Transactions on Aerospace and Electronic Systems, vol. 23, no. 3, May 1987, pp. 298-310.
- [15] A.S. Willsky, H.L. Jones. "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," IEEE Transactions on Automatic Control, vol. 21, no. 2, pp. 108-112, 1976.
- [16] Feili Hou, Feng Zhu, "Kalman filter for 3D motion estimation via lagrange interpolation and numerical integration," The 2004 IEEE International Conference on Robotics and Biomimetics, unpublished.
- [17] Y.K. Yang, Z.M. Zhu, Y.X. Huang, and J.S. Li., "Constructing estimate model for maneuvering target tracking by numerical differentiation: principle and applications," Chinese Journal of Electronics, vol. 30, no. 12, 2002, pp.1759-1762.
- [18] Y.K. Yang, Y.X. Huang, Z.M. Zhu, J.S. Li J., "Identification-compensation approach for unknown disturbance via the estimate model of numerical differentiation," Proceeding of the 4<sup>th</sup> World Congress on Intelligent Control and Automation, China, June 2002, pp.758-762.
- [19] J.C. Geromel, J.Bemussou, G.Garcia, M.C. de Oliveira, "H<sub>2</sub> and H<sub>∞</sub> robust filtering for discrete-time linear systems," SIAM Journal on Control and Optimization, vol. 38, no. 5, pp. 1353-1368, 2000.
- [20] Richard L. Burden, J. Douglas Faires, Numerical Analysis, US: Brooks Cole, 1985.
- [21] H.T. Kusofkf. The optimized filter and identification of control system. CN: National Defense industry Press, Beijing, 1984.
- [22] Y. Bar-Shalom, K.C. Chang. "Tracking a maneuvering target using input estimation versus the interacting multiple model algorithm," IEEE Transactions on Aerospace and Electronic Systems, vol. 25, no. 2, 1989, pp.296-299.
- [23] H.A.P. Blom, Y. Bar-Shalom. "The interacting multiple model algorithm for systems with markovian switching coefficients," IEEE Transactions on Automatic Control, vol. 33, no. 8, 1988, pp. 780-783.