

Data fusion for three-dimensional tracking using particle techniques

Huiying Chen

Youfu Li

City University of Hong Kong
Department of Manufacturing Engineering and
Engineering Management
83 Tat Chee Avenue
Kowloon, Hong Kong
E-mail: velvet.chen@student.cityu.edu.hk

Abstract. Robustness and tracking speed are two important indices for evaluating the performance of real-time 3-D tracking. We propose a new approach to fuse sensing data of the most current observation into a 3-D visual tracker with particle techniques. With the proposed data fusion method, the importance density function in the particle filter can be designed to represent posterior states by particle crowds in a better way. This makes the tracking system more robust to noise and outliers. On the other hand, because particle interpretation is performed in a much more efficient fashion, the number of particles used in tracking is greatly reduced, which improves the real-time performance of the system. Simulation and experimental results verified the effectiveness of the proposed method. © 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.2835013]

Subject terms: 3-D tracking; particle filter; data fusion; importance density.

Paper 070461R received May 31, 2007; revised manuscript received Aug. 6, 2007; accepted for publication Aug. 8, 2007; published online Jan. 23, 2008.

1 Introduction

The task of three-dimensional (3-D) visual tracking is to continuously estimate and update the position and orientation of a target in 3-D space.¹ As a state-space estimation problem, 3-D tracking can be modeled with the aid of parametric models. However, due to the varying degrees of uncertainty inherent in system modeling and the complexity of the system noise, the visual system is often subject to elements of non-Gaussianity, nonlinearity, and high dimensionality, which usually preclude analytic solution. It is a common strong belief that the problem of state measurement ultimately remains best handled within the framework of statistical inference. Instead of using linearization techniques, the estimation problem is solved directly with Bayesian methodology.² However, the Bayesian paradigm involves calculation of high-order integrals for the time state estimation. Thus, in the last few decades, many approximate filtering schemes, which are well known as methods of *particle filtering* (PF), and also as condensation or sequential Monte Carlo (SMC) methods,³⁻⁵ have been developed to seek a simulation-based way to surmount the problems.

The particle techniques provide a very feasible, convenient, and parallelizable approach to the estimation of posterior system states in visual tracking. However, the performance of PF depends much on the number of particles used, and the number of particles is usually determined by the dimension and structure of the state space. A typical 3-D tracking problem with six degrees of freedom often requires thousands of particles,⁶ which can run foul of computational complexity and further interfere with real-time performance for tracking agile motion.

On the other hand, in generic particle techniques, the most common strategy is to sample from the probabilistic

model of the states' evolution,^{3,7-10} i.e., to choose the transition prior $P(x_k|x_{k-1})$ as the importance density function. As a result, the proposed density function used for sampling is independent of the most current measurement, and thus it is sensitive to outliers and may decrease the reliability of the tracking algorithm. In order to tackle this problem, various approaches have been employed using variational optimization, Gaussian approximation, or local linearization methods.

Isard and Blake¹¹ developed a method, called ICON-DENSATION, to integrate both high-level and low-level information into a particle filter by importance sampling. Gordon et al.³ presented a prior editing method, in which an ad hoc acceptance test was used to propose particles in regions of high likelihood. They designed and computed the residual error as the criterion for particle sample rejection. The problem with this approach is that it is too heuristic and can be computationally intensive unless the rejection rate is small. Some approaches^{12,13} have used the extended Kalman filter (EKF) to calculate the approximation to the true posterior as the importance density. Although the EKF moves the importance density towards the likelihood, this method may introduce error due to linearization. van der Merwe et al.¹⁴ improved the method by employing more accurate approximations from the unscented Kalman filter (UKF) instead.

Though the preceding approaches enhanced the robustness of the PF algorithm, they did not provide a way to relieve the computational burden. Lu et al.¹⁵ addressed a method to solve the importance density problem and reduce the number of particles at the same time. They developed a RANSAC-based approach to modify the importance density. Nevertheless, this method is still considered computationally complex in our 3-D tracking context.

In this paper, we propose a new approach to fuse sensing data of the most current observation into a 3-D visual tracker with particle techniques. Incorporating the sensing

data, the importance density function is modified so that it can interpret the posterior particles in a better way, and therefore the particle crowd can represent the posterior states in a much more efficient fashion. Consequently, the number of particles used in tracking is greatly reduced. We have proved that about 100 particles are enough to perform a real-time 3-D tracking process with an active vision system.

In Sec. 2, we present the particle structure. The task of 3-D tracking using an active vision system is modeled in Sec. 3. We then describe the data fusion method and its algorithm for optimizing the importance density in Sec. 4. Subsequently, in Sec. 5, simulation and experimental results are given to show the effectiveness and superiority of the proposed method.

2 Framework of the Particle Filter

At the k 'th state, let the unknown state vector be x_k . Let the probability function of the prior state vector be represented by $P(x_k|x_{k-1})$. Then assume that we have k independent observations of the state vector, $\mathbf{y}=(y_1, y_2, \dots, y_k)$, drawn from a normal population with unknown mean and known variance. Given the sample information of the prior states, the likelihood function can be defined as

$$P(\mathbf{y}|x_k) \sim N(\mu_l, \sigma_l^2). \tag{1}$$

According to Bayes's rule, the posterior probability function of the state vector can be calculated using the following equation:

$$P(x_k|\mathbf{y}) = \frac{P(x_k|x_{k-1})P(\mathbf{y}|x_k)}{P(\mathbf{y})}. \tag{2}$$

The value of $P(\mathbf{y})$, which is the reciprocal of the normalizing constant for the probability density function (pdf) in Eq. (2), can be calculated as $P(\mathbf{y})=\int P(x_k)P(\mathbf{y}|x_k) dx_k$. Because $P(\mathbf{y})$ is a constant, Eq. (2) can be written as

$$P(x_k|\mathbf{y}) \propto P(x_k|x_{k-1})P(\mathbf{y}|x_k). \tag{3}$$

The weight of particle i can be defined from the likelihood function as

$$w_k^i \propto w_{k-1}^i \frac{P(x_k|x_{k-1})P(\mathbf{y}|x_k)}{q(x_k|x_{k-1}, \mathbf{y})}, \tag{4}$$

where $q(x_k|x_{k-1}, \mathbf{y})$ is the importance density function. The particle weights can be normalized as

$$w_k^i = \frac{w_k^i}{\sum_{i=1}^N w_k^i}, \tag{5}$$

where N is the total number of particles. Consequently, the approximation of the posterior density is expressed as

$$P(x_k|\mathbf{y}) \approx \sum_{i=1}^N w_k^i \delta(x_k - x_k^i), \tag{6}$$

with δ a Dirac delta function.

3 Task Statement with an Active Vision System

The 3-D tracking task here is to be performed with an active vision system¹⁶ using pattern projection, which is similar to a passive stereo vision system with one of the cameras replaced by a projector. Using a color-encoded structured light pattern,¹⁷ the active vision system can yield good results in 3-D visual sensing with a single view. Comparing to other 3-D sensing systems, such as the laser scanner or stereo vision, the active vision system has the merit that it can find feature points (points on the projected pattern) easily, even when the target object has a smooth and continuous surface.

3.1 State Transition Model

Feature points on the structured light pattern in an active vision system are denoted by $\mathbf{p}_i^{\text{patt}}=(x_i^{\text{patt}}, y_i^{\text{patt}}, z_i^{\text{patt}})^T$ on a certain light plane.¹⁸ Here $\mathbf{p}_i^{\text{patt}} \in S_{\text{patt}}$, where S_{patt} is a point set of the light pattern. Suppose that the target object is rigid. Then the target object, with a structured light pattern projected on it, can be defined as a set of 3-D feature points in the object reference frame. We denote these feature points by $\mathbf{P}^{\text{obj}}=\{\mathbf{p}_i^{\text{obj}}\}_{i=1}^{N_o}$, where $\mathbf{p}_i^{\text{obj}}=(x_i^{\text{obj}}, y_i^{\text{obj}}, z_i^{\text{obj}})^T$. If the target object's model is known according to the system configuration parameters, then $\mathbf{p}_i^{\text{obj}}$ can be obtained using $\mathbf{p}_i^{\text{patt}}$.

When the position and orientation of the object are tracked by the active vision system, we define at time k the state vector of the object as $x_k=(\mathbf{T}_k, \mathbf{\Omega}_k)^T$, with $\mathbf{T}_k=(X_k, Y_k, Z_k)$ the position of the object frame in a fixed world frame, and $\mathbf{\Omega}_k=(\phi_k, \theta_k, \psi_k)$ the orientation of the object frame in the fixed world frame.

Let $\mathbf{P}=\{\mathbf{p}_i\}_{i=1}^N=(x_i, y_i, z_i)$ be the set of object feature points in the fixed world frame. Then \mathbf{P} can be obtained from \mathbf{P}^{obj} by the following transformation:

$$\mathbf{P} = \mathbf{R}(\mathbf{\Omega})\mathbf{P}^{\text{obj}} + \mathbf{T} = \mathbf{R}(\psi)\mathbf{R}(\theta)\mathbf{R}(\phi)\mathbf{P}^{\text{obj}} + \mathbf{T}, \tag{7}$$

where $\mathbf{R}(\psi)$, $\mathbf{R}(\theta)$, $\mathbf{R}(\phi)$ and are the rotation matrices corresponding to roll, pitch, and yaw.

In the transition of the tracking state, the velocities can be modeled as an independent first-order Gaussian random walk model with the following definition:

$$(\dot{\mathbf{T}}_{k+1}, \dot{\mathbf{\Omega}}_{k+1})^T \sim N((\dot{\mathbf{T}}_k, \dot{\mathbf{\Omega}}_k)^T, \mathbf{\Sigma}_V), \tag{8}$$

where $N(\mu, \sigma)$ is a Gaussian distribution with mean μ and standard deviation σ , and $\mathbf{\Sigma}_V$ is a diagonal matrix comprising the variances of velocities, $\mathbf{\Sigma}_V = \text{diag}(\sigma_{\dot{x}}, \sigma_{\dot{y}}, \sigma_{\dot{z}}, \sigma_{\dot{\phi}}, \sigma_{\dot{\theta}}, \sigma_{\dot{\psi}})$. Using the velocities, the object position and orientation can be computed by

$$\mathbf{\Omega}_{k+1} = \mathbf{\Omega}_k + \Delta t \cdot \dot{\mathbf{\Omega}}_k, \tag{9}$$

$$\mathbf{T}_{k+1} = \mathbf{T}_k + \Delta t \cdot \mathbf{R}(\mathbf{\Omega}_k) \cdot \dot{\mathbf{T}}_k, \tag{10}$$

where Δt is a short sampling time. The prior $P(x_k|x_{k-1})$ can be obtained from Eqs. (8)–(10).

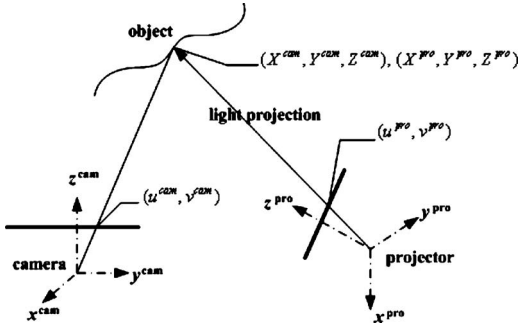


Fig. 1 Coordinates in the active vision system.

3.2 Observation Model

Figure 1 illustrates the representation of point coordinates in the active vision system.¹⁹ For the camera, the relationship between 3-D coordinates of an object point from the view of the camera, $\mathbf{X}^{\text{cam}} = (X^{\text{cam}} \ Y^{\text{cam}} \ Z^{\text{cam}} \ 1)^T$, and its projection on the image, $\mathbf{m}^{\text{cam}} = (\lambda u^{\text{cam}} \ \lambda v^{\text{cam}} \ \lambda)^T$, is given by

$$\mathbf{m}^{\text{cam}} = \mathbf{M}^{\text{cam}} \mathbf{X}^{\text{cam}}, \quad (11)$$

where \mathbf{M}^{cam} is a 3×4 perspective matrix of the camera. Obviously, \mathbf{X}^{cam} is a function of the state vector x_k . Equation (11) expresses the direct measurement by the camera. We term this process *passive sensing*. Such data are a kind of midway results in the active visual system.

Similarly, the projector is regarded as a pseudocamera in that it casts an image rather than detects it. The relationship between 3-D coordinates of the object point from the view of the projector, $\mathbf{X}^{\text{pro}} = (X^{\text{pro}} \ Y^{\text{pro}} \ Z^{\text{pro}} \ 1)^T$, and its back-projection on the projector sensor (LCD or DMD), $\mathbf{m}^{\text{pro}} = (\kappa u^{\text{pro}} \ \kappa)^T$, is

$$\mathbf{m}^{\text{pro}} = \mathbf{M}^{\text{pro}} \mathbf{X}^{\text{pro}}, \quad (12)$$

where \mathbf{M}^{pro} is a 2×4 inverse perspective matrix with the form

$$\mathbf{M}^{\text{pro}} = \begin{pmatrix} k^{\text{pro}} & 0 & u_0^{\text{pro}} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}_{2 \times 4}. \quad (13)$$

The relationship between the camera view and the projector view is given by

$$\mathbf{X}^{\text{pro}} = \mathbf{R}_t \mathbf{X}^{\text{cam}} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{X}^{\text{cam}}, \quad (14)$$

in which \mathbf{R}_t is a 4×4 matrix standing for three-axis rotation and translation. Substituting Eq. (14) into (12) gives

$$\mathbf{m}^{\text{pro}} = \mathbf{M}^{\text{pro}} \mathbf{R}_t \mathbf{X}^{\text{cam}}. \quad (15)$$

Let

$$\mathbf{H} = \mathbf{M}^{\text{pro}} \mathbf{R}_t = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{pmatrix}_{2 \times 4}, \quad (16)$$

where \mathbf{r}_1 and \mathbf{r}_2 are 1×4 row vectors. Equation (15) becomes

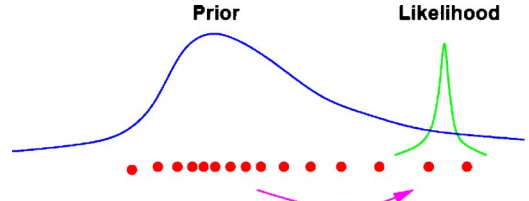


Fig. 2 Degeneracy of the importance density interpretation (from Ref. 14).

$$(u_0^{\text{pro}} \mathbf{r}_2 - \mathbf{r}_1) \mathbf{X}^{\text{cam}} = 0. \quad (17)$$

Combining Eqs. (11) and (17) gives

$$\begin{pmatrix} \mathbf{M}^{\text{cam}} \\ u_0^{\text{pro}} \mathbf{r}_2 - \mathbf{r}_1 \end{pmatrix} \mathbf{X}^{\text{cam}} = \begin{pmatrix} \mathbf{m}^{\text{cam}} \\ 0 \end{pmatrix}, \quad (18)$$

or

$$\mathbf{Q} \mathbf{X}^{\text{cam}} = \mathbf{m}_+^{\text{cam}}, \quad (19)$$

where \mathbf{Q} is a 4×4 matrix. Then the observation model of the 3-D target location from active sensing can be determined by

$$\mathbf{m}_+^{\text{cam}}(y_k) = \mathbf{Q} \mathbf{X}^{\text{cam}}(x_k). \quad (20)$$

4 Data Fusion for Importance Density Optimization

4.1 Motivation

In the particle framework, the posterior density is approximated by a weighted sum of N particles that are drawn from an *importance density*. Thus, the choice of the importance density directly influences the property of sampling, and it is a crucial factor. If the importance density is too small for some part of the support state of the posterior, one will need a very large sample size to approximate the state properly. On the contrary, if the importance density is well designed, only a small number of particles will be needed to draw a good representation of the state.

In generic particle techniques, the transition prior $P(x_k | x_{k-1})$ has often been employed as the importance density. This is a reasonable choice. Firstly, because $P(x_k | x_{k-1})$ is handy and easy to calculate, it will not increase the computational complexity of the system. Secondly, according to continuity, the distribution of the state x_k may not differ much from that of its prior state x_{k-1} in a short-time propagation; thus $P(x_k | x_{k-1})$ can represent the posterior state intuitively. However, because $P(x_k | x_{k-1})$ is independent of the most current observation information, it is sensitive to outliers. For instance, if the new measurements appear in the tail of the prior or if the likelihood is too peaked in comparison with the prior¹⁴ (see Fig. 2), $P(x_k | x_{k-1})$ will fail to interpret the posterior properly.

Here we propose a method to fuse passive sensing data into the importance density function for optimizing particle sampling to reduce the number of particles required. There are two reasons why we use passive sensing data rather than the active sensing data for reference. Firstly, active

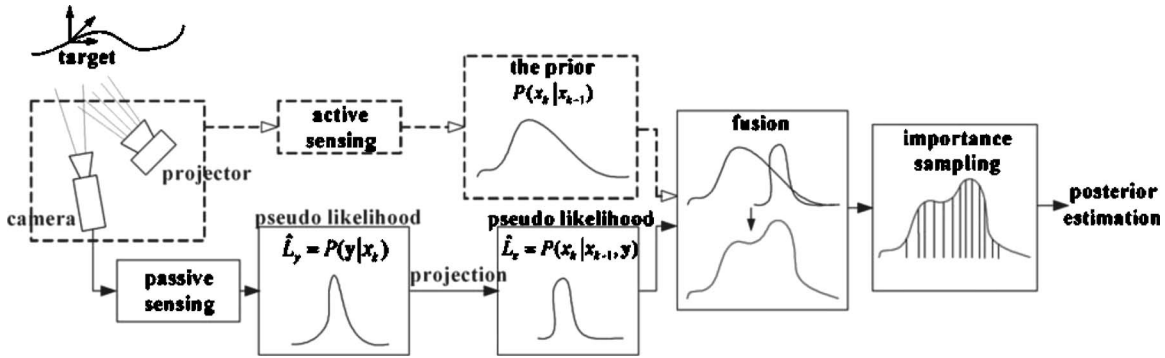


Fig. 3 Data fusion with updated passive sensing data.

sensing adds the projection part (of the projector) into the triangulation for feature acquisition, which is more complex in calculation and will increase the computational cost if we fuse it in every iteration of particle operation. Secondly, the errors of active sensing include distortion and projection errors of the projector, while in passive sensing, for simplicity's sake, we may only consider the sensing error of the camera, which is comparatively easy to model.

4.2 Data Fusion Algorithm

In order to obtain a better expression for the posterior state, the importance density function should be moved towards the region of high likelihood. However, because the supporting variables are different, the likelihood function cannot be used directly to modify the importance density.

Our approach to surmounting this problem is that we first generate a *pseudolikelihood* function with the latest passive sensing data and then project the pseudolikelihood into the importance density space and modify the importance density by fusing the sensing data in it. The pseudolikelihood function is generated with the most current observation of certain reference feature points through passive sensing. It is a subset of the likelihood function, and it can represent the likelihood function to a certain extent. The advantage of the pseudolikelihood function is that it can be projected into the importance density space easily by using the inverse procedure of the passive sensing observation

model (a monocular camera model). The basic idea of our approach is illustrated in Fig. 3.

Some reference feature points should be selected before we perform data fusion. These are geometical feature points on the target (such as corners and centers), and they are different from feature points on the structured light pattern for active sensing. The reference points should be easy to detect during the tracking process. On target objects with geometical primitives, those points are relatively easy to find. Considering the possibility of occlusion, it is better to choose more than one reference point. Nevertheless, according to our simulation study, increasing the number of reference points did not yield significant improvement of system performance.

Suppose, in a set of reference points, the observation of passive sensing (of a monocular camera) can be expressed as a function of the current state with noise

$$y_{k_j}^R = g_j(x_k, \zeta), \quad j = 1, \dots, N_R, \quad (21)$$

where ζ is the noise and N_R is the number of reference points. Equation (21) can be looked on as a pseudolikelihood function

$$y_k^R \sim \hat{L}_y = P(y|x_k^R). \quad (22)$$

On the contrary, the current state can be estimated using the inverse function of Eq. (21) as

Table 1 Passive data fusion algorithm in the particle filter.

Assume that at the previous state $k-1$, we have the particle crowd $\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^N$. Then proceed as follows at time k :

Sampling: Simulate $x_k^i \sim P(x_k|x_{k-1})$.

Pseudolikelihood computation: Simulate $\tilde{x}_k^i \sim \hat{L}_x = P(x_k|x_{k-1}, y_k^R, \zeta)$.

Data fusion: Let $x_k^i = \alpha x_k^i + (1-\alpha)\tilde{x}_k^i$, with α a data fusion factor, $0 \leq \alpha \leq 1$; draw samples from the new importance density.

Updating of weights: Compute the weights according to the likelihood function, and conduct normalization.

Resampling: Replace $\{x_k^i, w_k^i\}^N$ by $\{x_k^i, 1/N\}^N$.

Table 2 Comparison of the algorithm running times.

Filter	Running time (s)
Our PF	0.0150
EKF	0.0146
GPF (200 particles)	0.0919
GPF (400 particles)	0.1232
GPF (800 particles)	0.1477

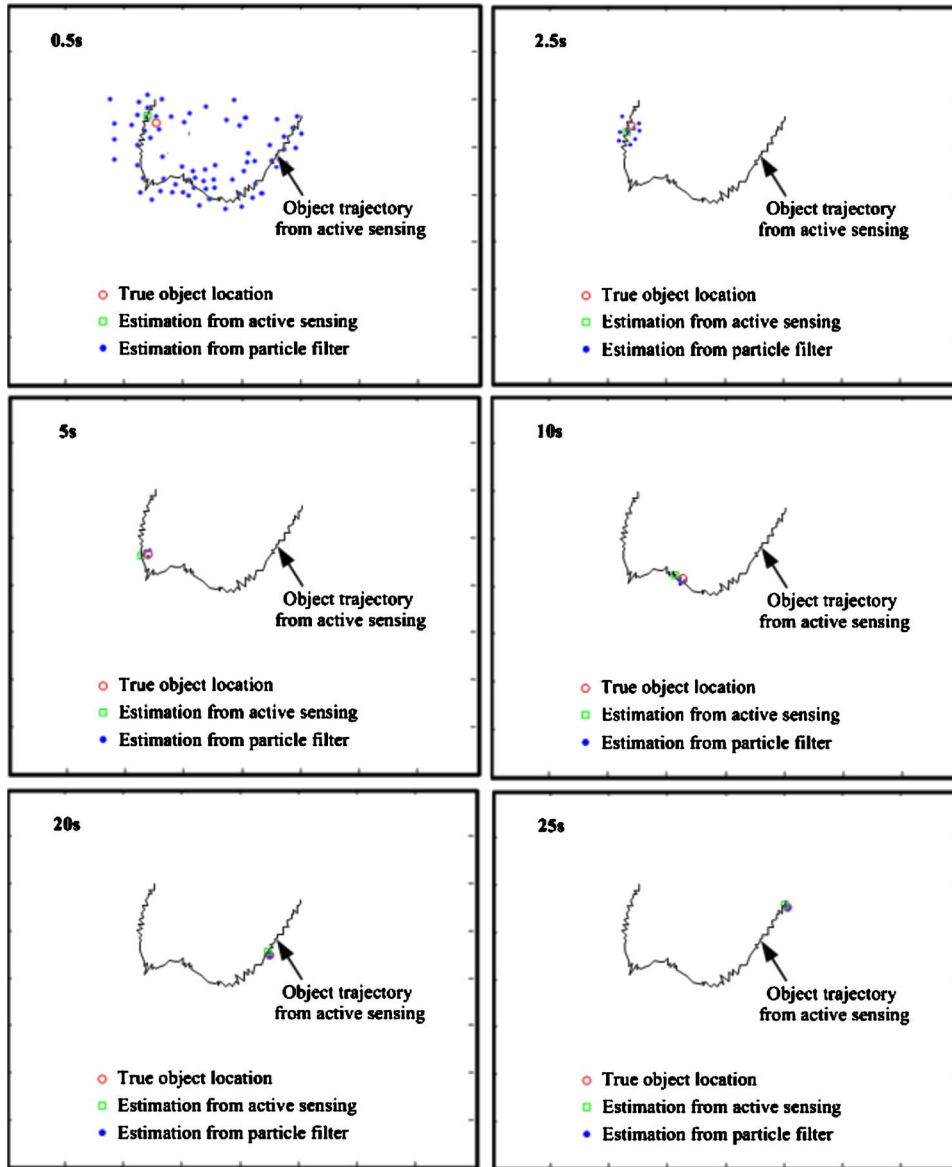


Fig. 4 Active visual tracking without data fusion.

$$x_k = g^{-1}(x_{k-1}, y_k^R, \zeta), \tag{23}$$

which is in fact a projection of the pseudolikelihood into the importance density (x_k) space,

$$\hat{L}_x = P(x_k | x_{k-1}, y_k^R, \zeta). \tag{24}$$

Then Eq. (24) can be used to achieve the algorithm for data fusion as shown in Table 1.

4.3 Importance Density Optimization

Degeneracy is a common problem with particle filters (see Ref. 12), especially when the importance density function is not well designed. As a result of degeneracy, all but one particle will have negligible weight after a few state transitions. Degeneracy implies a waste of computational resources in that a large effort is engaged to update particles whose contribution to the approximation of posterior states

is almost zero. Doucet¹² has shown that the variance of the importance weights can only increase over time, so that degeneracy is an inevitable phenomenon with general sequential importance sampling schemes.

We can here adopt the *effective sampling size*,^{20,21} which is a suitable measure of degeneracy, as a criterion to guide the data fusion process for optimizing the importance density function. According to Refs. 20 and 21, the effective sampling size N_k^{eff} at state k is defined as

$$N_k^{\text{eff}} = \frac{N_s}{1 + \text{Var}(w_k^*)}, \tag{25}$$

where w_k^* is taken as the *true weight* indicated in Eq. (4), and N_s is the number of samples. Because N_k^{eff} cannot be evaluated exactly,²² an estimate \hat{N}_k^{eff} of N_k^{eff} is calculated as

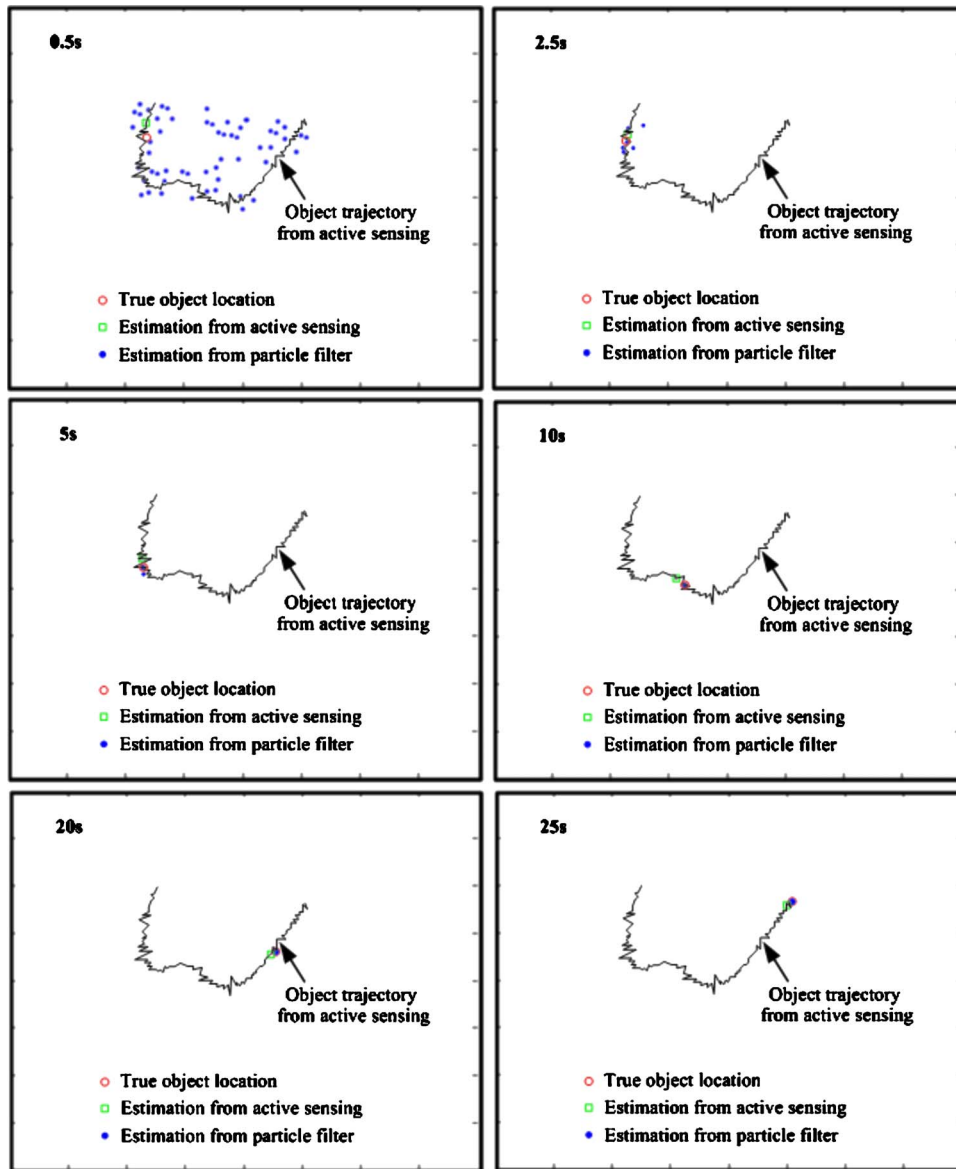


Fig. 5 Active visual tracking with passive data fusion.

$$\hat{N}_k^{\text{eff}} = \frac{1}{\sum_{i=1}^{N_s} (w_k^i)^2}, \quad (26)$$

where w_k^i is the normalized weight indicated in Eq. (5).

A large \hat{N}_k^{eff} implies that the likelihood is located close to the prior $P(x_k|x_{k-1})$, so that the particle crowd from the prior x_k^i can be relied on better. Thus, \hat{N}_k^{eff} can be used to evaluate the data fusion factor α (see Table 1) for importance density optimization by choosing

$$\alpha = \lambda \hat{N}_k^{\text{eff}} \quad (27)$$

with λ a positive scale factor and $0 \leq \alpha \leq 1$.

Then the new particle crowd after data fusion is

$$x_k^{*i} = \alpha x_k^i + (1 - \alpha) \hat{x}_k^i, \quad (28)$$

where x_k^i is drawn from the prior and \hat{x}_k^i is drawn from the pseudolikelihood function.

5 Simulation and Experimental Results

5.1 Simulation Results

5.1.1 Effectiveness of data fusion

We first studied and verified the effectiveness of passive data fusion. Figures 4 and 5 show the simulations of the proposed PF for active tracking with and without the fusion of passive information. In Fig. 4, without the fusion of passive sensing information, the particle tracker used about 16 s to converge to the true object location. Figure 5 shows that under the same condition, fusion of passive sensing

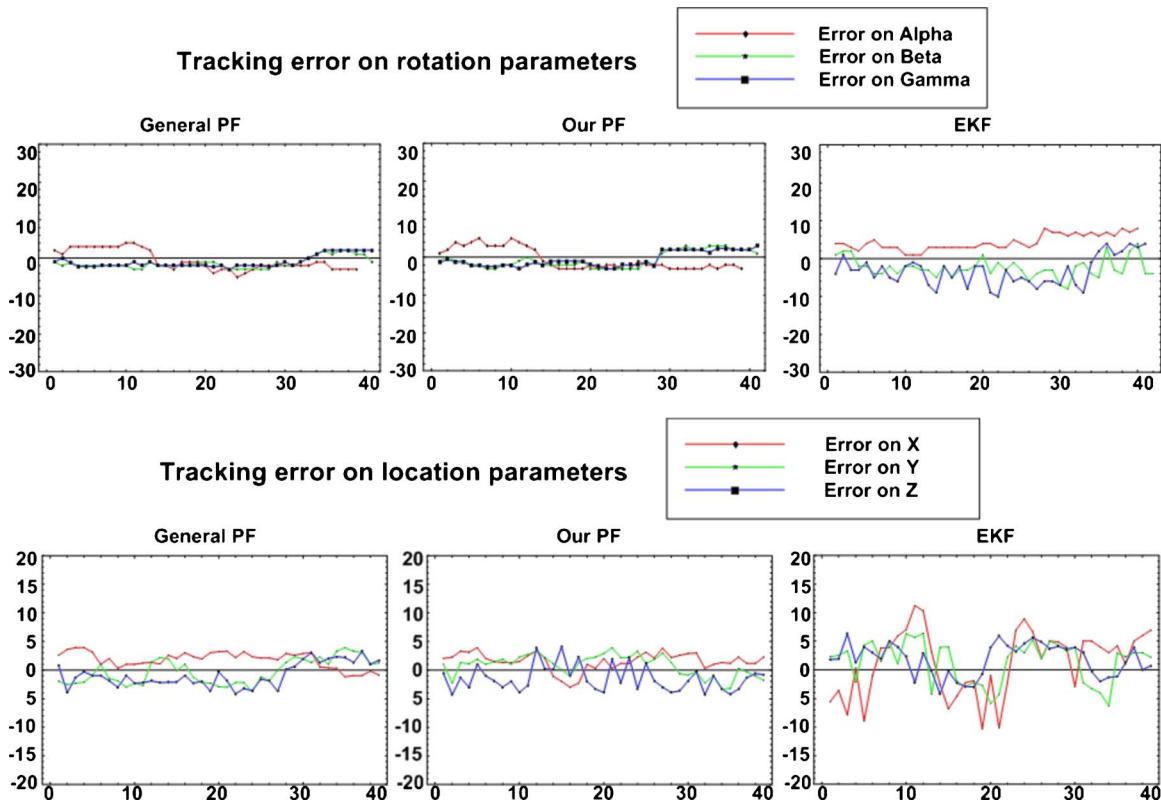


Fig. 6 Tracking accuracy comparison.

data helps the particle filter to achieve faster convergence. In this case, the active tracker took about 8 s to converge.

5.1.2 Tracking accuracy

Figure 6 shows the estimation error for active tracking by different methods. Here, the tracking error was defined as the distance between the estimated location of the tracking object and its true value. The generic PF, which employed 800 particles, performed the best, while the extended Kalman filter (EKF) performed the worst because of its disadvantage in dealing with multimodality. The proposed PF with data fusion, even with only 100 particles, achieved performance approximately as good as the generic PF.

5.1.3 Tracking speed

With a well-expressed importance density, the proposed PF can use only a small number of particles for accurate estimation. Therefore, it can achieve better real-time performance with expedition. Simulation results demonstrate the superiority of the proposed method in comparisons with EKF and generic PF (GPF).

Because the EKF does not involve calculations of sampling and resampling as generic PF does, it can achieve faster real-time performance than generic PF. Table 2 shows the algorithm running time for our PF, the EKF, and the generic PF with different numbers of particles. With only 100 particles, the proposed PF excels the GPF in running time.

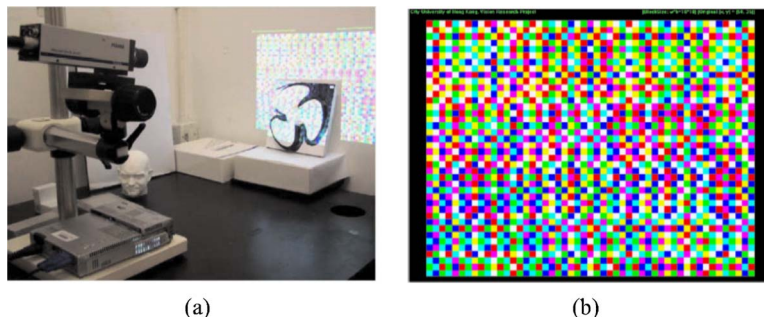


Fig. 7 Our active vision system using color-encoded structured light.

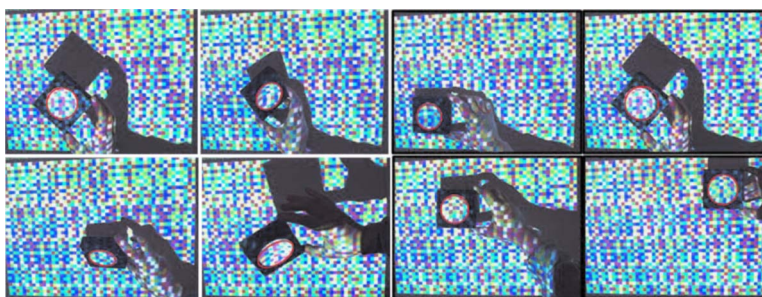


Fig. 8 Tracking a concave object with the proposed method.

5.2 Experiment

The proposed tracking method was implemented with an active vision system that consists of a PULNIX TMC-9700 CCD camera and a PLUS V131 DLP projector [as shown in Fig. 7(a)]. When the system is used in a visual tracking task, the projector projects color-encoded structured light [see Fig. 7(b)]¹⁸ onto the surface of the target object. Through observation and the use of triangulation, a time sequence of 3-D object positions and orientations can be obtained.

We used a concave object as the target, which can provide three degrees of freedom of translational motion and two of rotational motion in a 3-D space. A PF with 100 particles was employed. Some snapshots of the tracking are shown in Fig. 8. A tracking rate of about 12 ft/s was achieved in the implementation. As shown in Fig. 8, the tracking error was mainly affected by the quality of active sensing. For example, in the frame in the second row and second column of Fig. 8, larger tracking error is observed. This is because the target happened to move to a position such that the structured light pattern could not be detected clearly. A similar situation is shown in the frame in the first row and fourth column. When the relative depth to the projector was not large enough, the light pattern started to blur on the target surface, which influenced the sensing for target tracking.

6 Conclusion

In this paper, a new approach to fuse passive sensing data into an active 3-D visual tracker with particle techniques is presented. With the data fusion method, the importance density of particles can be modified to interpret posterior states in an efficient way. Therefore, the number of particles used in tracking is greatly reduced. Simulation and experimental results show that the proposed method improves the tracking accuracy as well as the tracking speed in comparison with other methods. Further research is underway in exploring the use of the fusion method for optimal tracking with a reconfigurable active vision system.

Acknowledgment

The work described in this paper was fully supported by a grant from the Research Grants Council of Hong Kong (Project No. CityU117605).

References

1. H. Y. Chen and Y. F. Li, "Enhanced 3D tracking using nonsingular constraints," *Opt. Eng.* **45**(10), 107202 (2006).
2. J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, Wiley (1994).
3. N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc. F, Radar Signal Process.* **140**(2), 107–113 (1993).
4. A. Doucet, N. D. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag (2001).
5. M. Isard and A. Blake, "CONDENSATION—conditional density propagation for visual tracking," *Int. J. Comput. Vis.* **29**(1), 2–28 (1998).
6. D. Fox, "KLD-sampling: adaptive particle filters," in *Proc. Conf. on Neural Information Processing Systems*, MIT Press (2001).
7. D. Avitzour, "A stochastic simulation Bayesian approach to multitarget tracking," *IEE Proc., Radar Sonar Navig.* **142**(2), 41–44 (1995).
8. E. R. Beadle and P. M. Djuric, "A fast weighted Bayesian bootstrap filter for nonlinear model state estimation," *IEEE Trans. Aerosp. Electron. Syst.* **33**(1), 338–343 (1997).
9. M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Proc. 4th Eur. Conf. on Computer Vision*, B. Boustou and R. Cipolla, Eds., pp. 343–356, Springer (1996).
10. G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *J. Comput. Graph. Stat.* **5**, 1–25 (1996).
11. M. Isard and A. Blake, "ICONDENSATION: unifying low-level and high-level tracking in a stochastic framework," in *Proc. 5th Eur. Conf. on Computer Vision*, H. Burkhardt and B. Neumann, Eds. pp. 893–908, Springer (1998).
12. A. Doucet, "On sequential simulation-based methods for Bayesian filtering," Technical Report CUED/F-INFENG/TR 310, Dept. of Engineering, Cambridge Univ. (1998).
13. M. K. Pitt and N. Shephard, "Filtering via simulation: auxiliary particle filters," *J. Am. Stat. Assoc.* **94**(446), 590–599 (1999).
14. R. van der Merwe, A. Doucet, N. de Freitas, and E. Wan, "The unscented particle filter," Technical Report CUED/F-INFENG/TR 380, Dept. of Engineering, Cambridge Univ., 2000.
15. L. Lu, X. T. Dai, and G. Hager, "A particle filter without dynamics for robust 3D face tracking," in *Proc. 2004 IEEE Conf. on Computer Vision and Pattern Recognition, Workshops* (2004).
16. P. Besl, "Active optical range imaging sensors," *Mach. Vision Appl.* **1**, 127–152 (1988).
17. P. M. Griffin, L. S. Narasimhan, and S. R. Yee, "Generation of uniquely encoded light patterns for range data acquisition," *Pattern Recogn.* **25**(6), 609–616 (1992).
18. B. Zhang, Y. F. Li, and Y. Wu, "Self-recalibration of a structured light system via plane-based homography," *Pattern Recogn.* **40**(4), 1368–1377 (2007).
19. Y. F. Li and S. Y. Chen, "Automatic recalibration of an active structured light vision system," *IEEE Trans. Rob. Autom.* **19**(2), 259–268 (2003).
20. N. Bergman, "Recursive Bayesian estimation: navigation and tracking applications," PhD Dissertation, Linkoping Univ., Linkoping, Sweden (1999).
21. J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamical systems," *J. Am. Stat. Assoc.* **93**, 1032–1044 (1998).
22. M. S. Arulampalam, M. Gordon and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.* **50**(2), 174–188 (2002).



Huiying Chen received the bachelor's degree in mechatronics engineering from South China University of Technology, China. She obtained the master's degree in precision machinery engineering from the University of Tokyo, Japan. She is currently a PhD candidate in the Department of Manufacturing Engineering and Engineering Management at City University of Hong Kong. Her research interests include robot

vision, visual tracking, and dynamic view planning.



Youfu Li received the BS and MS degrees in electrical engineering from Harbin Institute of Technology China. He obtained the PhD degree from the Robotics Research Group, Department of Engineering Sciences, of the University of Oxford in 1993. From 1993 to 1995 he was on the postdoctoral research staff in the Department of Computer Sciences, University of Wales, Aberystwyth, UK. He is currently an associate professor in the Department of Manufacturing Engineering and Engineering Management at City University of Hong Kong. His research interests include robot sensing, robot vision, sensor-based control, sensor-guided manipulation, 3-D vision, visual tracking, mechatronics, and automation. He is an associate editor of *IEEE Transactions on Automation Science and Engineering* (T-ASE).

robot vision, sensor-based control, sensor-guided manipulation, 3-D vision, visual tracking, mechatronics, and automation. He is an associate editor of *IEEE Transactions on Automation Science and Engineering* (T-ASE).