

Transactions of the Institute of Measurement and Control

<http://tim.sagepub.com/>

A hybrid approach to 3D arm motion tracking

Yaqin Tao and Huosheng Hu

Transactions of the Institute of Measurement and Control 2008 30: 259

DOI: 10.1177/0142331208090965

The online version of this article can be found at:

<http://tim.sagepub.com/content/30/3-4/259>

Published by:



<http://www.sagepublications.com>

On behalf of:



Institute of Measurement and Control

[The Institute of Measurement and Control](#)

Additional services and information for *Transactions of the Institute of Measurement and Control* can be found at:

Email Alerts: <http://tim.sagepub.com/cgi/alerts>

Subscriptions: <http://tim.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://tim.sagepub.com/content/30/3-4/259.refs.html>

>> [Version of Record](#) - Jul 25, 2008

[What is This?](#)

A hybrid approach to 3D arm motion tracking

Yaqin Tao and Huosheng Hu

Department of Computing and Electronic Systems, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

This paper presents a hybrid approach to 3D arm motion tracking for tele-rehabilitation applications. A particle filter (PF) algorithm is adopted in the proposed system to fuse data from inertial and visual sensors in a probabilistic manner. Multi-modal distributions of system states are propagated based on a ‘factor sampling’ technique. To avoid the problem of particle degeneracy in conventional PF algorithms, two strategies are adopted in our system, namely state-space pruning and an arm physical geometry constraint. Experimental results show that the proposed PF framework outperforms other fusion methods and tracking results are accurate in comparison to the ground truth provided by a commercial mark-based motion tracking system.

Key words: human motion tracking; inertial sensor; particle filters; tele-rehabilitation; visual tracking.

1. Introduction

Vision-based tracking and analysis of human motion has become an active research area because of its wide real-world applications, such as human computer interfaces (Gavrila, 1999), medical analysis (Moeslund and Granum, 2001) and surveillance monitoring (Gavrila, 1999). However, visual-only tracking systems suffer from motion blur and occlusion problems, which hamper many real-world applications. Therefore, a number of hybrid approaches using different sensors have been proposed into a motion tracking system in order to compensate for the sensors’ shortcomings and produce robust performance of a system, such as visual-audio (Wang *et al.*, 2003), visual-radar (Thomopoulos and Nillson, 1990) and visual-inertial (Strelow and Singh, 2004).

In this paper, visual and inertial sensors are deployed to track human arm motion for a tele-rehabilitation programme (Tao *et al.*, 2005). This hybrid tracking scheme is

Address for correspondence: Huosheng Hu, Department of Computing and Electronic Systems, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK. E-mail: hhu@essex.ac.uk

based on visual and inertial sensors in order to offer not only accurate motion tracking and stability, but also robust performance over fast motion and occlusions. It is intended to allow stroke patients to recover their motor ability in their home environments so that a burden of hospitals and physiotherapies can be relieved. However, its successful implementation relies on the proper use of fusion algorithms.

Recently, the particle filter (PF), also known as Sequential Monte Carlo Methods or Condensation (Isard and Blake, 1998a), has become popular in sensor fusion (Wang *et al.*, 2003), motion tracking (Isard and Blake, 1998b) and object recognition (Zhou *et al.*, 2002) in non-linear systems with multi-modal distributions. Its simplicity, flexibility and easy of implementation are realized by successfully modelling multi-model distributions based on sets of weighted particles. We employ the PF in this paper to carry out sensor fusion. There are varieties of versions of PFs (Arulampalam *et al.*, 2002; Gustafsson *et al.*, 2002). Apart from different names or different application areas, the essential difference is how they tackle degeneracy problems. There are three main factors that affect the performance of an PF, namely 1) the number of particles it uses; 2) particle impoverishment; 3) the choice of the proposal distribution. However, these three factors are not exclusively of each other. They are related, and a change in one factor will affect the others.

The number of particles defines how well the set of particles approximate the posterior distribution. In theory, when the number of particles becomes indefinite, the sample set will represent the true posterior probability density function (PDF) at each sampling time. So increasing the number of particles should improve fusion performance. However, computational cost increases exponentially with an increase in the number of particles. It is necessary to find a trade-off between PF performance and computational cost. Another problem that PF may suffer from is the particle impoverishment problem, which means the number of effective particles decreases as time passes, and particle sets do not reflect the true posterior density. Saving computational time and avoiding particle impoverishment are essentially the same problem. The principle is to use a small number of effective particles to represent the posterior distribution. Within the state-space, this means making a local instead of global search, ie, sampling particles near the most likely object positions. Different approaches have been proposed to deal with the particle degeneracy problem.

Some researchers work on how to sample particles effectively from a posterior distribution. Deutscher *et al.* (2000) proposed an annealed PF. The posterior PDF is implemented from coarse to fine by re-sampling particles around the peaks of a previous layer of a posterior PDF. No particles are sampled from the area of the state-space where the probability of system states is low, thus saving the number of particles required to model the posterior PDF. Sminchisescu and Triggs (2001) proposed a similar method called covariance scaled sampling, but it does not use a layered posterior PDF. Only the exact posterior PDF is used, and the method first finds the 'X' best local extremes in the posterior PDF, and later samples around the

extremes, so particles focus on the most likely regions and avoid low probability regions. The drawback is that they may fail to track sudden motion.

Employing an importance sampling algorithm is another way to improve PF performance. Isard *et al.* (1998b) combined PF with an importance sampling algorithm to fuse colour and contour features. It greatly reduced the number of samples required by employing a persistent image feature colour. The sampling procedure is first guided by using colour probability to locate the target object roughly, and then refined using contour tracking. Perez *et al.* fused different image features or information from other audio to track motion (Perez *et al.*, 2004). The colour features are also employed to serve there as the primary features, while sound or motion features are used as the auxiliary features to assist tracking. This method works well when primary features are consistently reliable.

Another method that can save the number of particles is Rao-Blackwellization (Doucet *et al.*, 1998). It divides the state-space into linear and non-linear sub-space and then uses PF and KF to propagate each sub-state-space separately. The main principle is to use as much linear property as possible. Experimental results show that it outperforms a conventional PF and uses a lesser number of particles (Gustafsson *et al.*, 2002). The basic idea behind this method is to reduce the size of state vectors to save on the number of particles, because the number of particles required increases with the size of the state vectors of a system.

We propose two strategies based on the arm physical geometry constraint and system configuration properties, to deal with the PF degeneracy and to improve PF performance. The first strategy is a state-space pruning method, which is used to reduce the number of particles. The second strategy is an arm physical constraint that can improve PF performance in narrow likelihood situation and reduce the particle impoverishment problem. We use the popular dynamic system model with constant acceleration as the proposal distribution in this paper. The experiment results prove that the proposed method outperforms other data fusion algorithms.

The rest of this paper is organized as follows. Section 2 describes arm motion tracking system configuration, and how the system configuration properties are used to improve PF performance by using a state-space pruning strategy as the first strategy. In Section 3, a human arm geometry constraint is employed as the second strategy to improve PF performance. Some experimental results are given in Section 4 to show the performance of the proposed hybrid system. Finally, conclusions and future work are presented in Section 5.

2. System overview and state-space pruning

2.1 System configuration

As shown in Figure 1, a video camera and an inertial sensor, MT9 from XSENS, are employed in our sensing and data fusion system to capture the arm motion

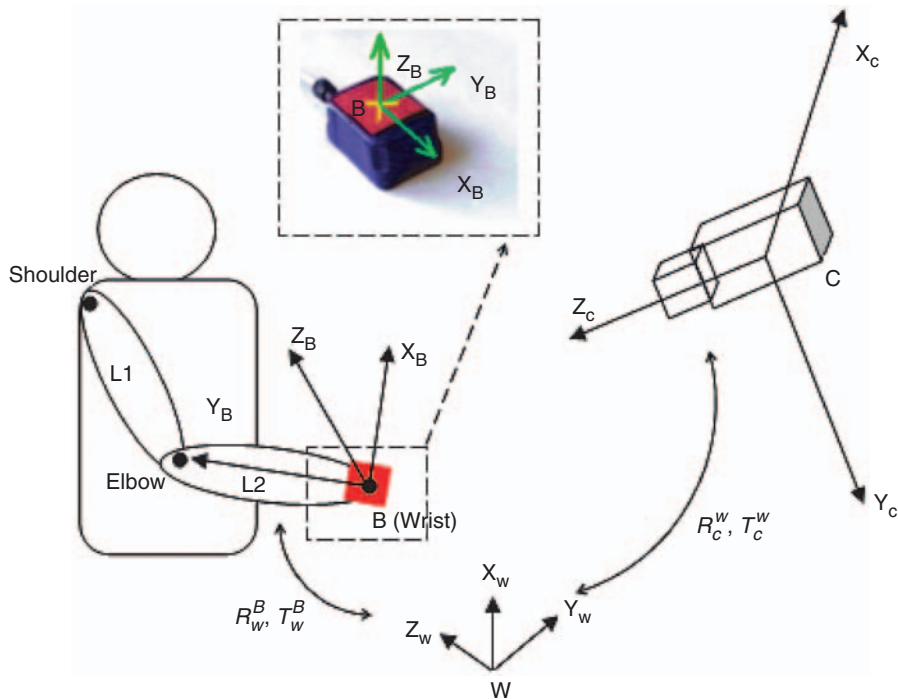


Figure 1 System configuration

of a subject. The video camera is fixed in the scene and used to capture the motion of an arm. An inertial sensor is attached to the wrist of the subject to measure his/her acceleration and orientation changes. Arm pose tracking is achieved by integrating visual and inertial information. Note that there are three co-ordinate systems involved in our arm motion tracking: 1) a camera frame (C); 2) an inertial sensor frame (B); and 3) a world co-ordinate system (W). Frame B is the body attached frame of the inertial sensor that is attached to the wrist joint of the human arm. Its origin coincides with the position of the wrist joint, and its y -axis is aligned with the forearm. The benefit of this configuration is demonstrated in the next section.

In order to simplify the tracking problem, one constraint is used in our method, ie, the shoulder joint is assumed fixed during the motion and the position is known as *a priori*. This is a realistic constraint because the shoulder joint normally moves much less as part of arm motion; such a constraint has been widely used in many arm motion tracking systems (Moeslund and Granum, 2000) and in rehabilitation (Rehabworks, 2007). We further assume that a hand is a stiff extension of the lower arm and the length of both forearm L_2 and upper arm L_1 are known *a priori*. The tracking task now becomes that of tracking elbow and wrist joints in 3D.

2.2 The state-space and pruning methods

Based on the above analysis, the pose of the arm can normally be represented using six variables in Cartesian co-ordinates, namely the three co-ordinates of the elbow position $P_{e,W}=(x_e, y_e, z_e)^T$ and the wrist position $P_{w,W}=(x_w, y_w, z_w)^T$, where W represents the world frame; e and w represent the elbow and wrist joints.

It is always desirable that the size of a state vector should be as small as possible, while capturing all the essential properties of the system. The computational time of an PF increases exponentially to the number of particles, while the number of particles grows with the size of system state. Therefore, by reducing the size of the state vector, it is possible to reduce computational cost of a filter. This can also improve the rate at which a search method converges into a global minimum or maximum in a small state-space.

Considering the property of the system configuration discussed in the last section and the property of a MT9 inertial sensor that measures not only acceleration but also angular velocity information, we could reduce the size of the state vector from six to three in Cartesian co-ordinates. The main idea is to convert the elbow and wrist joint tracking problem to track the wrist joint. The elbow joint is then inferred from wrist joint tracking and arm geometry information. Detailed procedures are as follows:

- 1) Since the origin of local frame B is defined at the wrist joint and the y -axis is aligned with the forearm, given the length of the forearm L_2 , the elbow joint in local frame B is constant and can be represented as:

$$P_{e,B} = (0, L_2, 0)^T \quad (1)$$

- 2) Given the pose R_W^B, T_W^B of the local frame B with respect to reference frame W , elbow joint position can be represented in reference frame W using Equation (2):

$$P_{e,W} = R_W^B * P_{e,B} + T_W^B = R_W^B * P_{e,B} + P_{w,W} \quad (2)$$

where T_W^B has the same meaning as $P_{w,W}$.

It is clear that once the orientation and position R_W^B, T_W^B of the wrist joint is calculated, it is straightforward to determine the elbow position by using Equation (2).

The MT9 inertial sensor contains a proprietary algorithm that can accurately calculate the absolute orientation of the moving sensor with respect to the reference frame in 3D space R_W^B (Figure 2). Our previous work in Tao *et al.* (2005) shows that the orientation information R_W^B from MT9 is quite accurate and has the potential to be used in the rehabilitation application to capture upper limb motion. We will use the orientation data from the inertial sensor directly in this paper and the unit quaternion $q=(q_w, q_1, q_2, q_3)$ to represent R_W^B . Quaternion representation facilitates co-ordinate transformation in a linear sensor, rather than in a non-linear form. Therefore, the 3D arm

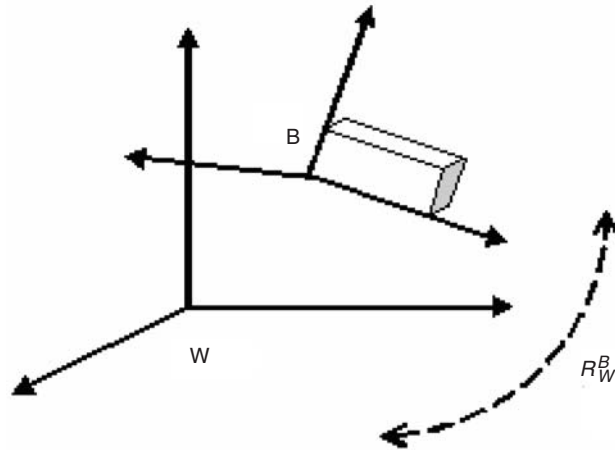


Figure 2 Frame B relative to reference frame W

motion tracking is converted to track the wrist joint position $P = P_{w,W} = (x_w, y_w, z_w)^T$. We use P to represent $P_{w,W}$ for simplicity in the rest of the paper.

3. Data fusion

3.1 System framework

In our arm motion tracking method, the inertial sensor is used to obtain the orientation and position of the wrist joint with respect to reference frame W . The visual sensor is used to track a 2D image projection of the wrist joint. Tracking results are then input to a fusion estimator PF to calculate the wrist joint position. Elbow position calculation is a straightforward task based on wrist joint calculation.

We employ a framework of PFs proposed by Gustafsson *et al.* (2002) for positioning, navigation and tracking:

$$x_{k+1} = Ax_k + B_u u_k + B_f f_k \quad (3)$$

$$z_k = h(x_k) + e_k \quad (4)$$

where x_k is the state vector, u_k is an input, f_k is process noise, z_k is a measurement and e_k is measurement noise. h is a non-linear function of the state vector x_k .

Equation (3) is a dynamical motion model of the system, which shows how system state evolves, given previous state and input information. A generic motion model with constant acceleration is assumed in wrist joint position tracking. The accelerations a_k^B and rate of turns ω_k^B of the wrist joint in inertial frame B are measurable from MT9. However, useful acceleration data is the wrist joint acceleration

in reference frame W , represented as a_k . It is calculated from inertial measurements using (5):

$$a_k = q_k * a_k^B * a_k' - g \quad (5)$$

where $*$ is the quaternion multiplication. g is gravity and q is the quaternion introduced in Section 2.2. a_k is then used as an input signal in the motion model:

$$x_{k+1} = \begin{bmatrix} P_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{pmatrix} I & I * T_S \\ 0 & I \end{pmatrix} x_k + \begin{pmatrix} T_S^2/2 * I \\ T_S \end{pmatrix} a_k + \begin{pmatrix} T_S^3/6 * I \\ T_S^2/2 \end{pmatrix} f_k \quad (6)$$

where T_S is sensor sampling time interval, and I is a 3×3 identity matrix. v_k is the first derivative of P_k and represents the velocity.

Equation (4) is a measurement model and represents the relationship between system state and measurement. It is derived from output of the visual sensor. The colour tracking method, Camshift (Bradski, 1998), is used to track the 2D position $z = (I_x, I_y, 1)^T$ of the wrist joint. Visual data is related to the state vector as follows:

$$z_k = \begin{bmatrix} I_x(k) \\ I_y(k) \\ 1 \end{bmatrix} = KI * (R_C^W * P_k + T_C^W) + e_k \quad (7)$$

where KI is a camera intrinsic parameter, and R_C^W, T_C^W is the pose of reference frame W represented in camera co-ordinate system C . They are both known *a priori*.

3.2 Fusion

The measurement model is formatted in probabilistic form as the observation density $p(z_k | x_k)$ in the following equation:

$$p(z_k | x_k) = p_{e_k}(z_k - h(x_k)) = p_{e_k}(z_k = KI * (R_C^W * P_k + T_C^W)) \quad (8)$$

where p_{e_k} is measurement noise density.

The motion model is formatted to density distribution $p(x_k | x_{k-1})$ as follows:

$$p(x_k | x_{k-1}) = p_{f_k}(B_f^+(x_{k+1} - Ax_k - B_u u_k)) \quad (9)$$

where B_f^+ is the Moore–Penrose pseudo-inverse, and p_{f_k} is the process noise density.

Having all the components ready, we can use the PF to fuse data recursively. Starting with a set of particles $\{s_{k-1}^{(i)}, \pi_{k-1}^{(i)}\}_{i=1}^N$, new particles $s_k^{(i)}$ are generated by

random sampling from the particle set and used for prediction by applying motion model $s_k^{(i)} = p(x_k | x_{k-1} = s_k^{(-i)})$. The particle's weight is updated according to measurement model $\pi_k^{(i)} = p(z_k | x_k = s_k^{(i)})$. The weights are normalized to sum to 1.

3.3 Layered sampling PF

During experiments, it was found that the PF fusion method introduced in the previous section did not give accurate results in comparison with the ground truth. After careful examination, we noticed that the PF suffers from severe particle impoverishment, since the measurement information is not sufficient. Only one visual feature point is employed from the visual sensor to update particle weights. It imposes fewer constraints on the subject's motion and the environment.

To design a more robust fusion algorithm, we introduce an arm physical geometry constraint into our fusion system, and allow it to serve as one of the measurement sources. This information is consistently available and does not place extra stress on image processing from the visual camera or subject motion. Experimental results show it improves the performance of the PF. The arm geometry we use is upper-arm and forearm lengths, which are assumed known *a priori* and constant during motion. This constraint is formulated as Equation (11) and served as a measurement equation to assist colour-based measurement.

$$L_2^2 = (P_{e,W} - P_{s,W})^2 + w_k = (R_W^B * P_{e,B} + P_k - P_{s,W}) + w_k \quad (10)$$

where $P_{s,w}$ is the position of the shoulder joint, assumed known and constant. P_k is wrist joint position at time k , and w_k is a noise variable.

Now the measurement equations include both colour tracking measurement Equation (7), denoted as $p(z_k^C | x_k)$ and geometry information Equation (10) $p(z_k^G | x_k)$, which are fused with motion model Equation (6) in a layered sampling framework described in the next section.

3.4 Implementation

Layered sampling PF implementation is described in this section. The particle set at a previous time is represented as $\{S_{k-1}^{(i)}, \pi_{k-1}^{(i)}\}_{i=1}^N$. Particles are propagated at time k as follows:

- 1) *Measurement update.* Compute the weights for each particle via the likelihood of colour tracking $\pi_k^{(i)} = p(z_k^C | x_k = s_k^{(i)})$ and normalize $\tilde{\pi}_k^{(i)} = \pi_k^{(i)} / \sum_{j=1}^N \pi_k^{(j)}$, $i = 1, \dots, N$.
- 2) *Re-sampling.* General a new set $\{s_k^{(i-)}\}_{i=1}^N$ by re-sampling using replacement N times from $\{s_{k-1}^{(j)}\}_{j=1}^N$ where $prob(s_k^{(i-)} = s_{k-1}^{(j)}) = \tilde{\pi}_k^{(j)}$, $j = 1, \dots, N$.
- 3) *Prediction.* Predict new particles $\{s_k^{(i+)}\}_{i=1}^N$ from $p(x_k | x_{k-1} = s_k^{(i-)})$, $i = 1, \dots, N$.
- 4) *Measurement update.* Compute the weights for each particle using the likelihood of arm geometry $\pi_k^{(i)} = p(z_k^G | x_k = s_k^{(i+)})$ and normalize $\tilde{\pi}_k^{(i)} = \pi_k^{(i)} / \sum_{j=1}^N \pi_k^{(j)}$, $i = 1, \dots, N$.

- 5) *Re-sampling.* General a new set $\{s_k^{(j)}\}_{j=1}^N$ by re-sampling using replacement N times from $\{s_k^{(i)}\}_{i=1}^N$ where $\text{prob}(s_k^{(i)} = s_k^{(j)}) = \tilde{\pi}_k^{(j)}$, $j = 1, \dots, N$.
- 6) Set $k = k + 1$ and iterate to Step 1.

In our previous research on the extended Kalman filter (EKF) (Tao and Hu, 2006), an arm geometry constraint was also employed to improve estimation accuracy, but it was used after EKF estimation and formulated into an optimization framework, which made the previous arm motion tracking method consisting of two separate frameworks: EKF and optimization. Here, the arm geometry constraint is used as one measurement and fused into the PF consistently.

4. Experimental results

4.1 Circle motion demonstration

We demonstrate the performance of our proposed method on an arm rehabilitation motion, circular motion, which follows a drawn circle on a table (Figure 3). The experiment set-up involves two tracking systems: our proposed system and the commercial marker-based system, Qualisys. Figure 3 shows a subject wearing both an inertial sensor and three ball markers.

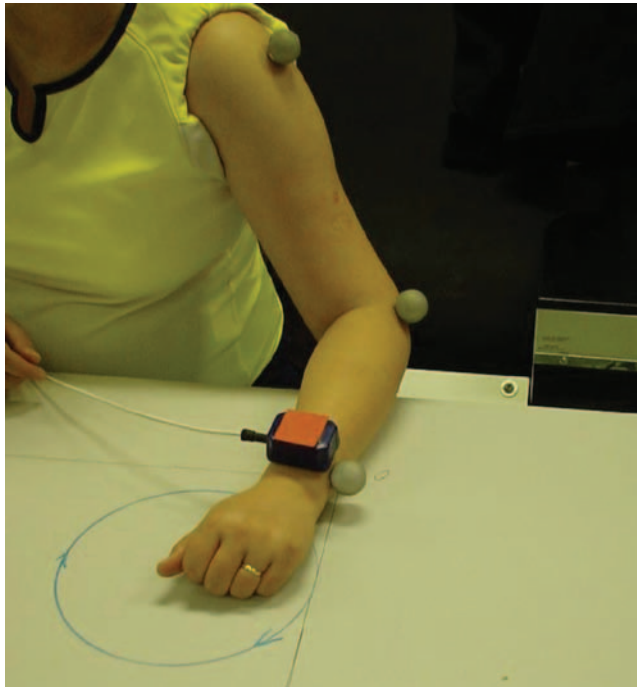


Figure 3 A subject performs a circle motion

uses three passive ball markers attached to the arm joints: shoulder, elbow and wrist. The markers are tracked by three special cameras to measure arm motion. These two systems operate at the same time to track the circle motion of a human arm. Tracking results from the marker-based system are regarded as a ground truth.

We fuse the inertial and visual data using different fusion algorithms, such as EKF, PF and the proposed PF with constraint. The fusion results from different fusion methods are compared with the ground truth. We illustrate the circle motion trajectories of the wrist joint in x , y and z co-ordinates from each algorithm respectively in Figure 4. The motion trajectories are represented in world co-ordinate system W .

As can be seen in Figure 4, the fusion results of a circle motion from different methods in x , y and z co-ordinates. The dashed lines represent results from EKF, the dotted lines are results from PF without constraint, the thin solid lines show results from PF with constraints and the bold solid lines show results from the marker-based system used as a reference. It is clear that the PFs with constraint match the ground truth well and give better results than the traditional EKF fusion algorithm and conventional PF, especially in y and z directions. The large deviations at end of all the traces are because of the drift problem of the inertial sensor. The inertial sensor is robust to fast motion, but suffers from drift problems on a long-term basis. Fusing visual and inertial data can improve the performance, but the tracking accuracy depends on the data fusion algorithms. Figure 4 clearly shows that our proposed fusion method gives the better results.

Figure 5 shows the error difference of each algorithm, compared with the ground truth. The two horizontal lines in each figure represent an error range of ± 0.05 m. It is very clear that the EKF algorithm produces the biggest error difference in all directions. The error differences are unbounded and reach 0.1 m in the x -axis, 0.6 m in the y -axis, and 0.4 m in the z -axis, at 80 frames. Obviously, this is not suitable for a rehabilitation application. The normal PF provides better results than the EKF algorithm; however, the error is still too large, especially in the y - and z -directions, which have an error difference of up to 0.4 and 0.2 m, respectively. In contrast, the error difference of the proposed PF accuracy is bounded to 0.05 m over 70 frames, as shown by the bold solid lines in Figure 5.

The fusion results of the circle motion from different methods are reconstructed in 3D and shown in Figure 6. It is clear from the figure that our proposed method gives the best results and is very close to the ground truth. Traditional PFs perform slightly better than the EKF method, but their accuracy is far from acceptable.

4.2 Fusion performance evaluated by statistical results

The statistical properties of average mean and standard deviation for the each method are calculated in our experiments to further evaluate the proposed fusion method. As shown in Table 1, it is clear that the inertial only tracking method suffers severe drift error. Error mean and deviation is as great as 21.31 cm/37.70 cm in the x -co-ordinate.

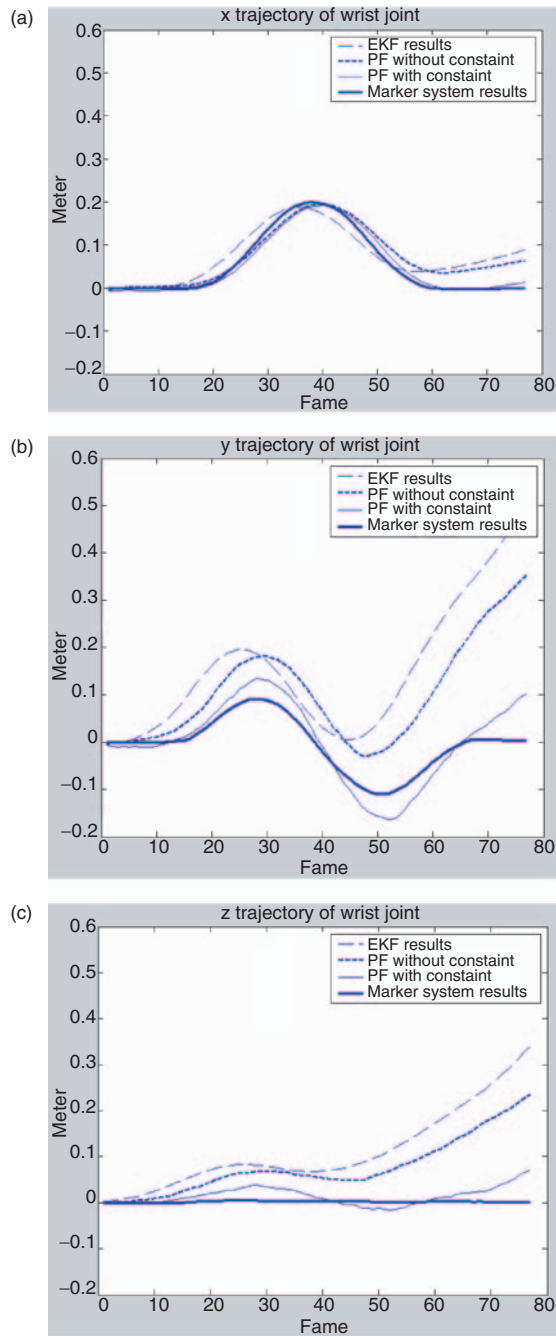


Figure 4 Fusion results of a circle motion using different methods: (a) *x*-motion trajectory of the wrist joint; (b) *y*-motion trajectory of the wrist joint; (c) *z*-motion trajectory of the wrist joint

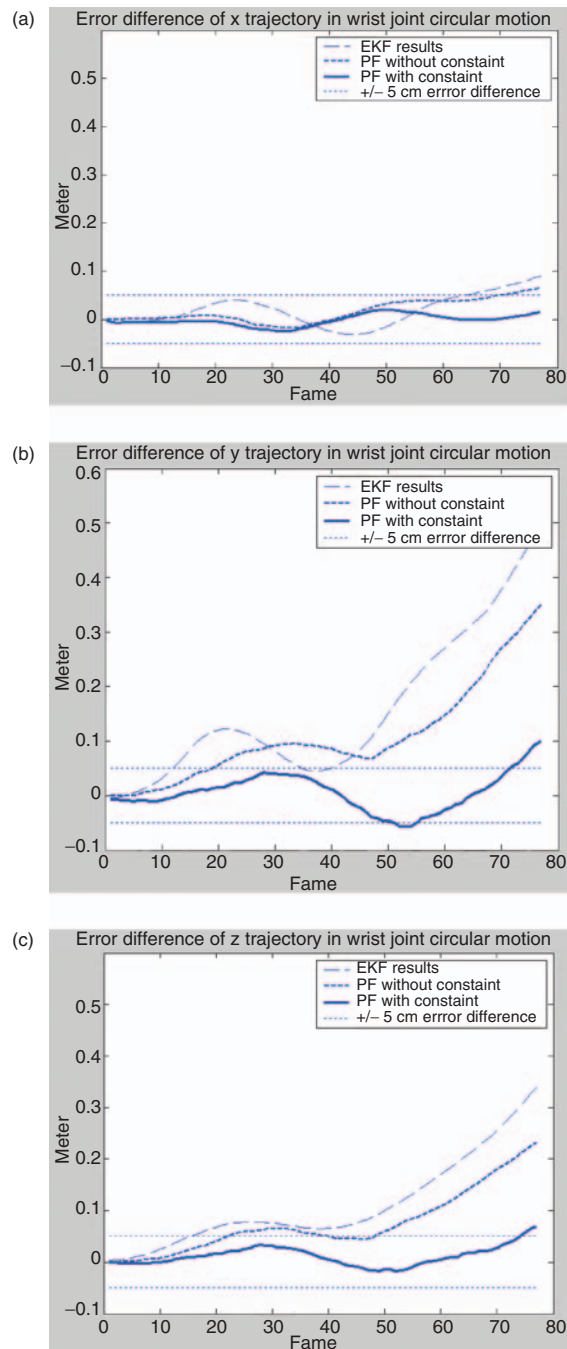


Figure 5 Error difference of different fusion methods compared with the ground truth: (a) error difference in the x -axis; (b) error difference in the y -axis; (c) error difference in the z -axis

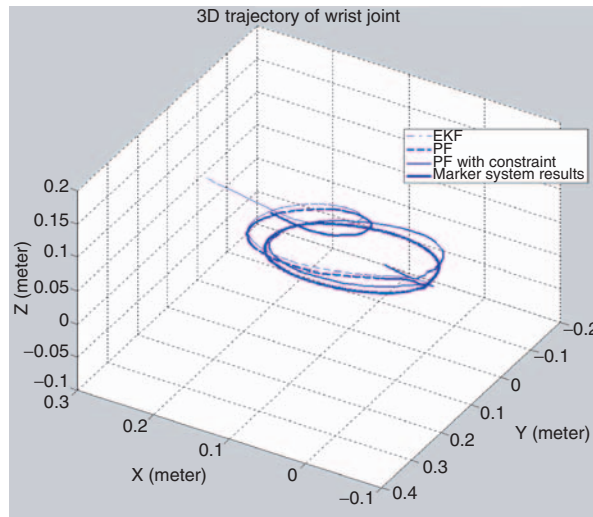


Figure 6 3D Reconstruction of a circle motion

Table 1 Mean error and standard deviation of each axis for a circle motion when using different algorithms

Circle motion	x-co-ordinate error (cm)	y-co-ordinate error (cm)	z-co-ordinate error (cm)
Inertial only	21.31/37.70	-12.41/16	6.74/5.23
EKF	0.4/3.30	10.56/9.39	7.83/5.72
PF	1.8/2.67	10.08/6.55	6.84/4.15
PF with constraint	-0.68/2.0	1.14/5.47	1.77/3.27

The EKF greatly improved the tracking results by fusing inertial data with visual data, but accuracy is not good enough. The PF provides slightly better performance than the EKF method. The PF with constraint gives the best performance of all algorithms, and its tracking errors are mainly in a range of ± 5 cm when compared with the ground truth.

5. Conclusions and future work

A new arm motion tracking system for tele-rehabilitation applications is presented in this paper, based on the integration of both vision and inertial sensors. The standard methods for clinical rehabilitation are marker-based motion tracking systems, eg, CODA (Charnwood Dynamics, no date) and Qualisys (no date). However, apart from the difficulties of calibrating both cameras and markers, these systems are too

expensive for a daily deployment by stroke patients at home, and too complicated for physiotherapists to interpret the tracking results of patients' motion. Visual-based marker-free tracking systems have also attracted a lot of research interest; however, they are normally computationally expensive and have low tracking accuracy. The proposed hybrid tracking scheme in this paper offers not only accurate motion tracking and stability, but also robust performance over fast motion and occlusions. It also reaches a good balance point between cost, accuracy and processing time.

Inertial and visual data fusion is implemented using PFs. The PF degeneracy problem is attacked by proposing a state-space pruning strategy and introducing an arm physical geometry constraint. Experimental results show it outperforms other fusion methods under the same circumstances. Although inertial and visual sensors have been widely used to track rigid objects, our proposed system is able to track non-rigid human arm motion in 3D.

Our future work will be focused on two directions in order to make the system more applicable and robust. First, a more efficient noise compression method will be developed, since the inertial sensor attached to a deformable human body generates a lot of noise, thus the fusion performance deteriorates. Second, it is desirable to release the fixed shoulder constraint, and extend the tracking method to track the upper body or whole body. Some other image features such as edge, contour, and suitable subject geometry and kinematics models can be used to implement the motion tracking task.

References

- Arulampalam, S., Maskell, S., Gordon, N. and Clapp, T.** 2002: A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking. *IEEE Transactions of Signal Processing* 50, 174–88.
- Bradski, G.R.** 1998: Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal* 2, 1–15.
- Charnwood Dynamics Ltd.** Codamotion tracking system. <http://www.codamotion.com/> Accessed 26 March 2008.
- Deutscher, J., Blake, A. and Reid, I.** 2000: Articulated body motion capture by annealed particle filtering. *Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, 13–15 June.
- Doucet, A., Godsill, S. and Andrieu, C.** 1998: On sequential Monte Carlo sampling methods for Bayesian filtering. Technical Report CUED/F-INFENG/TR. 310, Cambridge University Department of Engineering.
- Gavrila, D.M.** 1999: The visual analysis of human movement: a survey. *Journal of Computer Vision and Image Understanding* 73, 82–98.
- Gustafsson, F., Gunnarsson, F., Bergman, N., Forsslund, U., Jansson, J., Karlsson, R. and Nordlund, P.J.** 2002: Particle filters for positioning, navigation and tracking. *IEEE Transactions on Signal Processing* 50, 425.
- Isard, M. and Blake, A.** 1998a: CONDENSATION—conditional density propagation for visual tracking. *International Journal of Computer Vision* 29, 5–28.
- Isard, M. and Blake, A.** 1998b: Condensation: unifying low-level and high level tracking in a stochastic framework. *Proceedings of the European Conference on Computer Vision* 893–908.
- Moeslund, T. and Granum, E.** 2001: A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding* 81, 231–68.

- Moeslund, T.B. and Granum, E.** 2000: 3D human pose estimation using 2D-data and an alternative phase space representation. *Workshop on Human Modelling, Analysis and Synthesis at CVPR*.
- Prez, P., Vermaak, J. and Blake, A.** 2004: Data fusion for visual tracking with particles. *Proceedings of the IEEE* 92, 495–513.
- Qualisys.** Qualisys motion capture systems. <http://www.qualisys.com/> Accessed 26 March 2008.
- Rehabworks,** <http://rehabworks.ksc.nasa.gov/education/protocols/basicshoulder.php>, Accessed January 2007.
- Sminchisescu, C. and Triggs, B.** 2001: Covariance scaled sampling for monocular 3D body tracking. *Computer Vision and Pattern Recognition*, Kauai Marriott, Hawaii, 9–14 December.
- Strelow, D. and Singh, S.** 2004: Motion estimation from image and inertial measurements. *International Journal of Robotics Research* 23, 1157–95.
- Tao, Y. and Hu, H.** 2006: 3D arm motion tracking for home-based rehabilitation. *Proceedings of the 3rd Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT 06)*, Cambridge, 10–12 April.
- Tao, Y., Hu, H. and Zhou, H.** 2005: Integration of vision and inertial sensors for home-based rehabilitation. *Proceedings of the 2nd Workshop on Integration of Vision and Inertial Sensors (InerVis 2005)*, during ICRA'05, 18 April.
- Thomopoulos, B.C. and Nilsson, L.** 1990: 3-D motion tracking using stereo camera and range radar. *Proceedings of the SPIE*, Volume 1260, 21–35; B. Girod, editor, *Sensing and reconstruction of three-dimensional objects and scenes*, January.
- Wang, L., Hu, W. and Tan, T.** 2003: Recent developments in human motion analysis. *Pattern Recognition* 36, 585–601.
- Zhou, S., Krueger, V. and Chellappa, R.** 2002: Face recognition from video: a CONDENSATION approach. *International Conference on Automatic Face and Gesture Recognition*, Washington D.C., 20–21 May.