Acta Technica 61, No. 4A/2016, 141–150

Research on hand movement recognition based on static image decomposition

JUN HE[1]

Abstract. Estimation of human hand movement state is studied, that is, estimate the global pose of the human hand and the state of each local joint angle from the visual observation. In this paper, the static image sequence acquired by the camera is used as the observation input. Based on the improved Particle Swarm Optimization (PSO) particle filter algorithm study the human hand movement recognition, realize the interaction between the hand movement and the sphere.

Key words. Image sequence, particle filtering, human hand movement.

1. Introduction

It is an important research topic to use computer vision to analyze and recognize the multijoint hand movement. The main goal of this research is to detect and recover the human motion from the sequence of video images, and then to describe and understand the human hand movement behavior on this basis [1–2]. The visual analysis of the human hand movement can be applied to many fields, such as robot teaching and learning, human-computer interaction, virtual reality or three-dimensional animation [3]. At the same time, the visual analysis of human hand movement is a very challenging research topic, which involves image processing, computer vision, computer graphics, pattern recognition and artificial intelligence and other disciplines. In the past few decades, researchers at home and abroad have carried on a great deal of research around this subject. However, due to the inherent complexity of the project, the research results so far cannot meet the requirements of real-time, accuracy and robustness, and there is a certain distance from the practical application.

At present, some relatively mature mobile analysis systems in the market most need people to wear sensors or optical markers on the hand [4]. The sensor system (such as a data glove) acquires motion data by placing a sensor on a tracked object,

[1]Guizhou Electronic Information Career Technical College, Guizhou, 550000, China

such systems can guarantee a certain accuracy, but the equipment is often expensive, require complex calibration steps, and it can impede the movement of the object being tracked. The optical marking system places the optical mark on the joint of the human hand, then these markers are tracked by the camera, but in the movement, blocking will cause some of the mark is not visible, therefore, the system needs manual post-processing of the captured data. In comparison, the unmarked motion tracking method of computer vision is a more natural, more applicable and non-contact method. With the rapid increase of computer hardware processing capacity and the popularization of digital video equipment, the unmarked visual analysis of manpower moving has attracted the attention of many researchers both at home and abroad because of its wide application prospect, which has gradually become a research focus in computer vision field.

2. Methods

2.1. Algorithm

One of the main problems of the standard particle filter algorithm is the use of state transition prior model $p\left(x_i\left|x_{i-1}\right.\right)$, which does not consider the latest observation z_i, as the important density function, so the importance of the particle sampling process is suboptimal. In the tracking process, the standard particle filter algorithm needs to collect a large number of samples in order to approximate the true posteriori probability density distribution of the system state. The sample set too small will produce sample poor phenomenon, reduce the estimation accuracy, and even lead to sample set divergence and estimate failure. The three-dimensional human motion tracking is a high-dimensional problem, in the premise of calculation, due to the extreme sparsity of particle sampling in high-dimensional space, it is impossible to use a finite number of particles to effectively express the true posteriori distribution of human hand-like disease, which can easily lead to tracking failure.

The Gaussian particle swarm optimization (PSO) algorithm is an improved PSO algorithm [5], which uses the Gaussian distribution to generate the velocity vector. Its convergence is better than that of the classical PSO algorithm. Its particle velocity and position update equation is as follows:

$$v_{k+1}^i = |\mathrm{randn}|\left(p_k^i - x_k^i\right) + |\mathrm{randn}|\left(g_k - x_k^i\right) \tag{1}$$

$$x_{k+1}^i = x_k^i + v_{k+1}^i \tag{2}$$

where |randn| is a positive Gaussian random number that can be generated by the absolute value of the standard Gaussian distribution $N(0,1)$.

Particle Swarm Optimization Particle Filter (PSOPF) algorithm [6] is an algorithm that integrates the PSO algorithm into PF and optimizes the sampling process of PF. In this paper, the PSOPF algorithm is applied to human motion tracking in high-dimensional state space. The Gaussian PSO algorithm is used to optimize the matching error function $E(z, x)$. In the inter-frame state transition process, the $t-1$th moment convergence of the individual optimal position $p^i_{t-1,K}$ is to initialize the particles at time t

$$x^i_{t,0} = p^i_{t-1,K} + r \tag{3}$$

where $r \sim N(0, \sum)$ is the mean multidimensional Gaussian noise and $\sum$ is the covariance matrix. Its diagonal elements are determined according to the maximum interface angle or displacement variation.

However, Gaussian PSO algorithm with the standard PSO algorithm by updating the global extreme and extreme values of the particles themselves can search for the optimal solution only when a better solution is encountered. Then, the extreme value will be updated, thus narrowing the search neighborhood of the particle and making it easy to converge. In the process of particle swarm pursuit of the best particles, as it is getting closer to the best particles, the speed is getting smaller and smaller, and the entire group gradually shows a strong convergence. Therefore, the algorithm is easy to fall into the local optimum in the middle and late stages of the search process, resulting in the premature convergence phenomenon. This problem is particularly acute for high-dimensional multimodal problems such as hand motion tracking. In this paper, two improved methods are used to overcome the premature convergence of the algorithm and improve the diversity of the particle to enhance the global search ability of the algorithm.

First, introduce the simulated annealing into Gaussian PSO; use the simulated annealing algorithm to improve the updating condition of the extreme value of the particle. It cannot only accept the optimal solution, but also accept the degenerate solution to expand the global searching range. It is given by the formula

$$p^i_{k+1} = \begin{cases} x^i_{k+1} & if, r^i \min\left(1, \exp\left(-\Delta D/T_{k+1}\right)\right), \\ p^i_k & \text{otherwise}, \end{cases} \tag{4}$$

where T_{k+1} is the annealing temperature of the $(k+1)$th iteration. The value of $T_{k+1} = \alpha T_k$, α being the cooling coefficient that lies in interval $(0, 1)$. In all the experiments described in this paper, the values of α are 0.8. Symbol ΔD is the particle distortion change defined as $\Delta D = 1/f\left(x^i_{k+1}\right) - 1/f\left(p^i_k\right)$, where $f\left(x^i_{k+1}\right)$ and $f\left(p^i_k\right)$ are the new particle fitness value and old best individual fitness value, respectively. In this paper $f(x) = p(z\,|x)$, that is, the fitness value $f(x)$ of the particle is calculated using the observation likelihood $p(z\,|x)$ defined by $p(z\,|x) \propto \exp\left(-\lambda_e \cdot E(z, x)\right)$. Finally, r^i is a uniformly distributed random number from the interval [0, 1].

Secondly, starting from the third iteration of Gaussian PSO, local randomization is performed on the corresponding dimension of the finger joint angle according to the method in [7]. More specifically, in each iteration process, after the particle position update calculation, the particle position is recalculated with a small probability, which is selected as uniform sampling in the corresponding dimension value range.

It is shown in the formula

$$x_k^{ij} = \begin{cases} U\left(x^{j,\min}, x^{j,\max}\right) & \text{if } \mathrm{rand}^{ij} < \mathrm{Pr} \\ x_k^{ij} & \text{otherwise} \end{cases} \tag{5}$$

Here, $x^{j,\min}$ and $x^{j,\max}$ are the minimum and maximum values defined by the jth dimension of the search space, respectively, while $U(x^{j,\min}, x^{j,\max})$ is a random number evenly distributed in interval $[x^{j,\min}, x^{j,\max}]$. Symbol $\mathrm{rand}^{ij} = U(0,1)$ is a random number evenly distributed over the interval [0, 1]. Finally, Pr is a probability wide value constant for the treatment of particles, in this experiment its value is 0.01.

In this paper, the modified PSOPF is used as the tracking algorithm to search the human hand and object state parameters in $32(26+6)$-dimensional state space. In the particle filter framework, the convergence of the particle set is accelerated by optimizing the matching error of the Gaussian PSO, at the same time, using simulated annealing method to avoid the premature convergence of the particle, but canceling the local optimization in the process of randomization steps. Although the localized randomization method is suitable for tracking a variety of flexible movements individually, it does not apply to the tracking of hand-to-object interaction (such as human hand-crawling objects).

In the initial stage, due to the lack of time domain continuity information, this paper uses manual and object in their respective calibration positions to manually initialize the tracking process. The specific algorithm steps are as follows:

- Initialization: From the prior distribution of $p\left(x_{h-o,0}\right)$ sampling N particles, the weights are $1/N$, which is expressed as $\left\{x_{h-o,0}^i, 1/N\right\}_{i=1}^N$.
- Particle state transfer: This process is an important sampling process. The particle initial position is acquired using (3). With the newest observing value, (1) and (2) are used to iteratively evolve the particle set. The iteration will drive the particle to the high likelihood region. The particle diversity and mature convergence is assured through (4).
- Weight update: The observation likelihood is used to update the particle weight $w_t^i \propto w_{t-1}^i p\left(z_t \left| x_{h-o}^i\right.\right)$ and weight normalization. The maximum post-calibration is utilized to output the system state estimation value.
- Resampling: To avoid the degradation of particle weights, the sample set $\left\{x_{h-o,t}^i, w_t^i\right\}_{i=1}^N$ is resampled according to the weights. The new equal-weight sample set is $\left\{x_{h-o,t}^i, 1/N\right\}_{i=1}^N$.
- Is the judgment over? If yes, exit this algorithm, otherwise go to step 2.

2.2. Relevant models

This paper focuses on the interactive process between human hand and sphere. The method used in this paper is also applicable to the tracking of human hand interaction with other objects.

After defining the state variable (hand-sphere variable x_{h-o}), the matching error is defined as

$$E\left(z, x_{h-o}\right) = \lambda_d E_d\left(z, x_{h-o}\right) + \lambda_s E_s\left(z, x_{h-o}\right) + \lambda_m E_m x_{h-o}\,. \qquad (6)$$

The above equation shows that the matching error with the current observation input z is determined by the human hand motion state and object motion state. For the state transition model of human hands and objects, the first-order linear model is adopted in this chapter, and the multidimensional Gaussian noise is used as the disturbance.

3. Results and discussion

This experiment system is developed on the Visual Studio 2010 platform using open source cross platform 3D graphics engine Open Scene Graph (OSG) 3.0.0. The three-dimensional human hand and object model with DOF degrees of freedom nodes are loaded into the OSG. The control the movement of hands and objects is realized through osgSim::DOFTransfrom. In addition, in each iteration, the depth image of human hand and object model is generated by using Frame Buffer Object (FBO) in OSG, which is used to calculate the particle matching error and observation likelihood value in the algorithm.

In this paper, the experiments on the synthetic and real sequences are performed to verify the effectiveness of the hand-object tracking method. In all experiments, the improved PSOPF tracking method used in this chapter works with 60 particles, and for each image input, iterations are optimized 40 times. Experiments were run on an ordinary PU machine with dual-core Core 2 2.0 GHz CPU, 2.0GB of RAM and Nvidia GeForce 8400M GS GPU.

3.1. Synthesis of sequence experiments: Relevant models

Since it is not possible to obtain the real pose data directly from the sequence of images captured by the camera, this chapter uses a synthetic sequence with realistic pose data to further evaluate the algorithm quantitatively. The real value data for the synthesis sequence is the tracking result of the tracking system in the real sequence experiment, and the synthetic sequence of the human hand and the sphere is generated by the tracking result of the real sequence. The synthetic sequence is used as the input of the tracking system, and then for carrying out the experiments. Figures 1–7 show the results of tracking method proposed in this paper for a part of the tracking parameters and corresponding value of the comparison. The solid line depict the tracking results, while the dotted lines show the real values.

Figures 1-7 show the recognition results of human hand from a video frame by frame. The human hand captured in the video is permanently interactive with a sphere object. The sphere object works as an interference in the human hand recognition. When the palm moves along axis x, the actual movement and the recognition results converge perfectly in Fig. 1, indicating that the method can keep a high degree of consistency with the real situation. In Fig. 2, the actual movement is compared with the recognition results as the palm moves along axis y. Figure 2 also shows very good convergence of the actual results and tracking results. Figures 3–5 show the recognition results of PIP (proximal interphalangeal), MCP (metacarpophalangeal), TM (trapeziometacarpal) joint movement measured in angles.

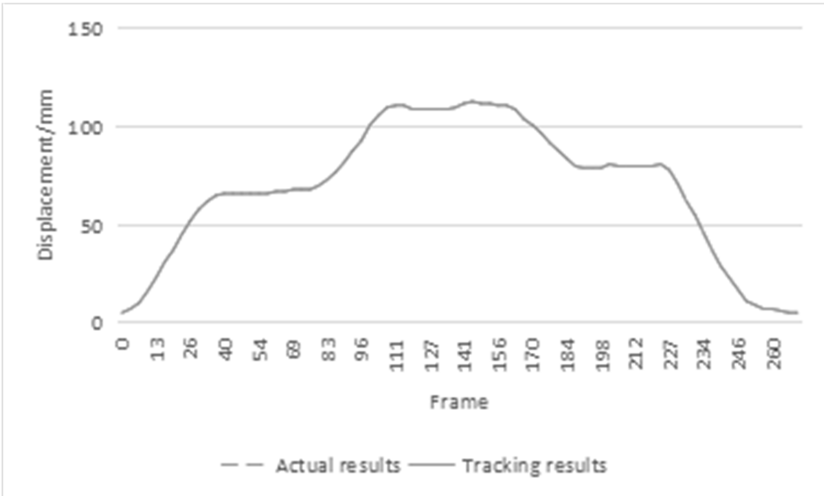

Fig. 1. Comparison of tracking results and real values of synthetic sequences between human hand and sphere: palm displacement along axis x

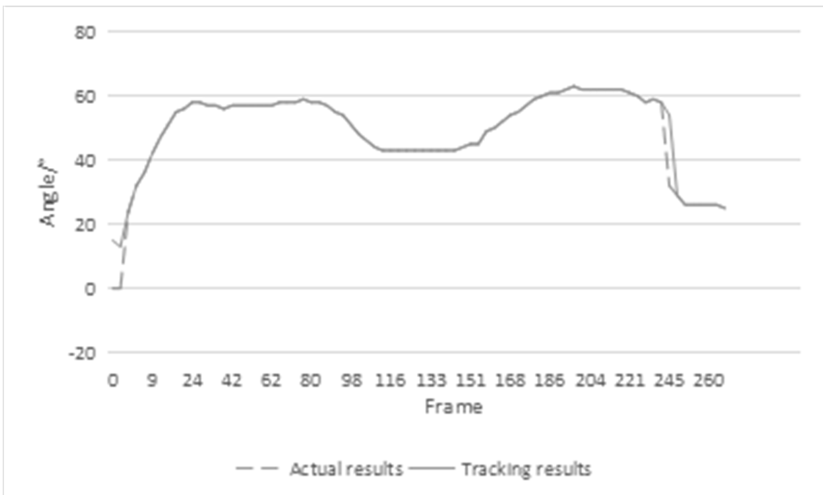

Fig. 2. Comparison of tracking results and real values of synthetic sequences between human hand and sphere: palm displacement along axis y

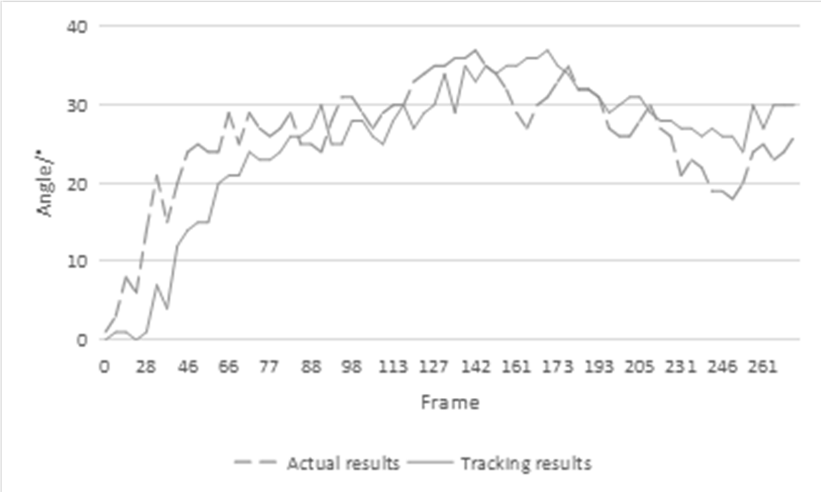

Fig. 3. Comparison of tracking results and real values of synthetic sequences between human hand and sphere: angle of ring finger PIP joint flexion

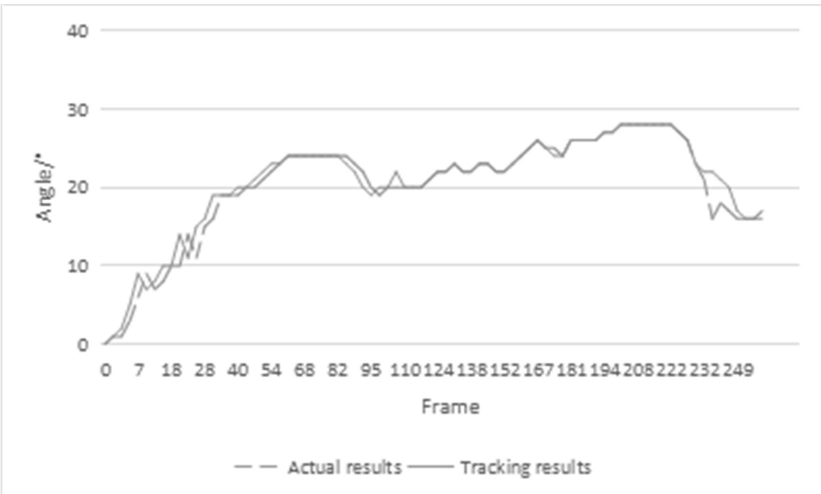

Fig. 4. Comparison of tracking results and real values of synthetic sequences between human hand and sphere: angle of thumb MCP joint flexion

4. Conclusion

Aiming at the difficulty of particle sampling in three-dimensional space, this paper integrates the swarm intelligence optimization method into particle filter, and using its powerful global optimization ability to improve the distribution of parti-

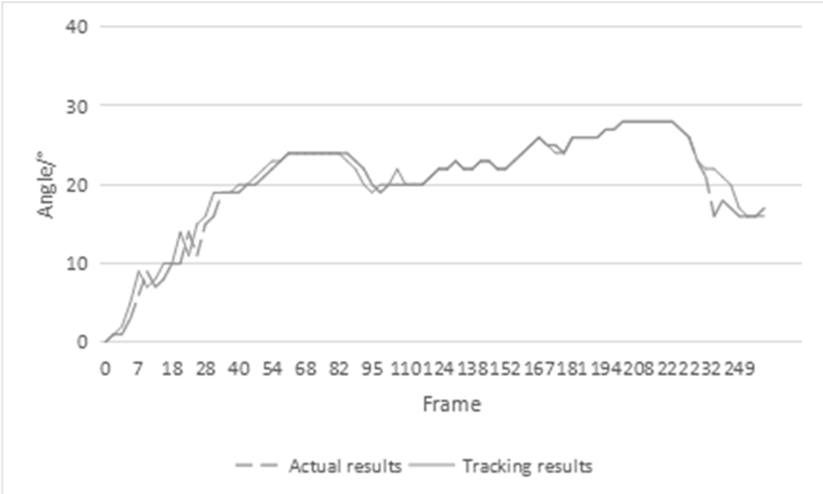

Fig. 5. Comparison of tracking results and real values of synthetic sequences between human hand and sphere: angle of thumb TM joint flexion

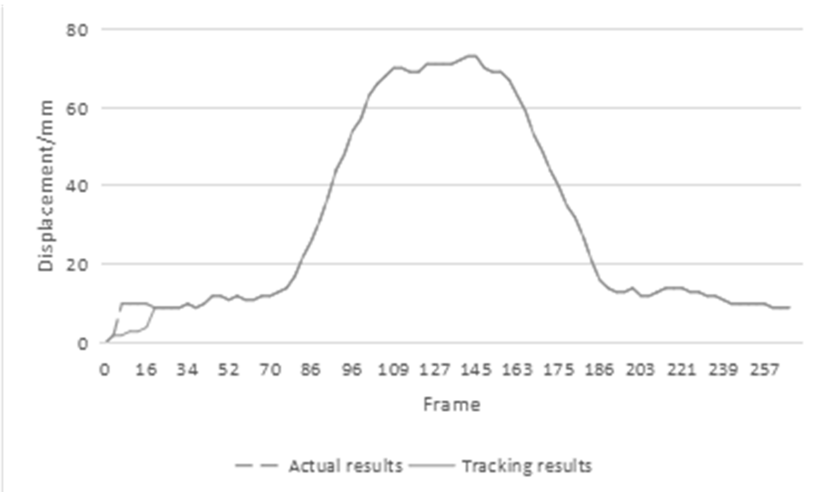

Fig. 6. Comparison of tracking results and real values of synthetic sequences between human hand and sphere: object displacement along axis x

cle filter samples, it proposes a specific three-dimensional hand tracking algorithm. This algorithm applies an existing Particle Swarm Optimization (PSO) particle filter algorithm to hand motion tracking in high-dimensional space, and in order to solve the problem of premature convergence in high dimensional-space, it uses simulated annealing and local randomization to improve the convergence of the algorithm. The improved algorithm can track the free movement of human hands and interaction between human and object.

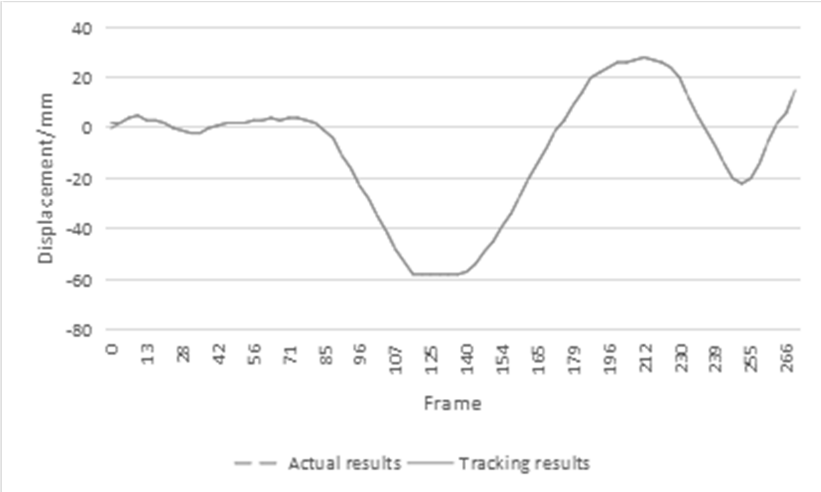

Fig. 7. Comparison of tracking results and real values of synthetic sequences between human hand and sphere: object displacement along axis y

Table 1 summarizes the results of the statistical analysis of the attitude parameters error on the whole sequence.

Table 1. Tracking error of synthetic sequences between human hand and sphere

Attitude parameter	Error mean	Standard deviation
Palm displacement along x	0.3453	0.2896
Palm displacement along y	0.6655	1.1481
Angle of ring finger PIP joint flexion	3.5181	2.5534
Angle of thumb MCP joint flexion	2.2100	1.8216
Angle of thumb TM joint flexion	0.7944	0.8104
Object displacement along x	0.6265	1.3522
Object displacement along y	0.3121	0.5192

References

[1] M. Protter, M. Elad: *Image sequence denoising via sparse and redundant representations.* IEEE Trans Image Process *18* (2009), No. 1, 27–35.

[2] E. Kalogerakis, O. Vesselova, J. Hays, A. A. Efros, A. Hertzmann: *Image sequence geolocation with human travel priors.* Proc. IEEE IC Computer Vision (ICCV), 29 Sept.–2 Oct. 2009, Kyoto, Japan, 253–260.

[3] M. Linderman: *EMG measured during controlled hand movement for biometric analysis, medical diagnosis and related analysis.* Patent US 8280503 B2, 2012.

[4] B. J. Galvan, A. S. Marco, J. F. Rolin, L. Delauney: *NeXOS contribution to the adaptation of system analysis engineering tools for mature and reliable ocean sensors.* Proc. IEEE Sensor Systems for a Changing Ocean (SSCO), 13–17 Oct. 2014, Brest, France, 1–6.

[5] X. Zhang, P. Wang, J. C. Xing, Q. L. Yang: *Particle swarm optimization algorithms with decreasing inertia weight based on Gaussian function.* Application Research of Computers *29* (2012), No. 10, 3710–3712.

[6] T. Krzeszowski, B. Kwolek, K. Wojciechowski: *Articulated body motion tracking by combined particle swarm optimization and particle filtering.* Lecture Notes in Computer Science *6374* (2010), 147–154.

[7] T. Yasuda, K. Ohkura, Y. Matsumura: *Extended PSO with partial randomization for large scale multimodal problems.* Proc. World Automation Congress (WAC), 19–23 Sept. 2010, Kobe, Japan, 1–6.

Received November 16, 2016