Vis Comput (2010) 26: 607–617
DOI 10.1007/s00371-010-0452-z

ORIGINAL ARTICLE

3D-freehand-pose initialization based on operator's cognitive behavioral models

Zhiquan Feng · Minming Zhang · Zhigeng Pan · Bo Yang · Tao Xu · Haokui Tang · Yi Li

Published online: 17 April 2010

Abstract Tracking, recognition and interaction based on 3D freehand are a part of our virtual assembly system, in which monocular camera is used to input online freehand videos and the hand pose tracker requires a reliable initial pose in the first frame. A novel approach to initializing 3D pose and position of freehand is put forward in this paper visualization of 3D hand model and modeling the operators' cognitive behaviors. Our approach is composed of three phases: hand posture recognition, coarse-tuning and fine-tuning. The operator moves his/her hand onto the to meet the needs of our virtual assembly system. The main contribution of this paper is that the three core techniques are for the first time integrated together, including human–computer interaction (HCI) in the process of initializing, projection of the 3D hand model in the period of coarse-tuning time. Then, the computer repeatedly fine-tunes the 3D hand model until the projection of the 3D hand model is completely superimposed onto the operator's hand image. We focus on exploring and modeling cognitive behavior of operator's hand upon which we design our initialization algorithm. Our research shows that cognitive behavioral models are not only beneficial to reducing cognitive loads for operators, because it makes the computers cater for the changes of the operators' hand poses, but also helpful to address high dimensionality of articulated 3D hand model. Our experimental results also show that the approach presented in this paper is easier, more pleasurable and satisfactory experience for the operators. Our initialization system has successfully been applied to our 3D freehand tracking system and a simulation virtual assembly system.

Keywords 3D Freehand pose model · Features extraction · Initialization · Visualization · Cognitive behavioral models

Z. Feng (✉)
School of Information Science and Engineering, University of Jinan, Shandong 250022, P.R. China
e-mail: ise_fengzq@ujn.edu.cn

M. Zhang · Z. Pan
State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058, P.R. China

M. Zhang
e-mail: zmm@cad.zju.edu.cn

Z. Pan
e-mail: zhigengpan@gmail.com

B. Yang · T. Xu · H. Tang · Y. Li
Provincial Key Laboratory for Network Based Intelligent Computing, University of Jinan, Shandong 250022, P.R. China

B. Yang
e-mail: yangbo@ujn.edu.cn

T. Xu
e-mail: ise_xut@ujn.edu.cn

H. Tang
e-mail: ise_tanghk@ujn.edu.cn

Y. Li
e-mail: liyi@ujn.edu.cn

1 Introduction

There has been a great emphasis lately in Human Computer Interaction (HCI) research to create easier use of interfaces by directly employing natural communication and manipulation skills for humans. Adopting direct sensing in HCI will allow the deployment of a wide range of applications in more sophisticated computing environments, such as Virtual Environments (VEs) or Augmented Reality (AR) systems. The development of these systems involves addressing challenging research problems including effective input/output techniques, interaction styles and evaluation methods. In the input domain, the direct sensing approach requires capturing and interpreting the motion of a hand. The hand is the most effective, general-purpose interaction tool due to its dexterous functionality in communication and manipulation [1].

One of the keys to track freehand by Monte Carlo method is to acquire initial state of a hand pose with some accuracy [1–4]. Especially in recursive and model-based freehand tracking approaches, initialization of 3D-freehand-pose is a necessary step, yet it is a troublesome problem because in most of the studies, initial state of a hand pose is manually complemented.

Motivated by the application of human–computer interface in AR and VE, bare hand tracking requires a reliable initial pose in the first frame. There are many difficulties in the initialization. Firstly, it remains challenging to recover 3D structure from a single hand image in computer vision domain. Secondly, freehand is a typical articulated object with high dimensionality, and to find out the real 3D hand model from nearly unlimited hand gestures is almost impossible unless approximate methods are used, just as most of the single-frame pose estimation approaches do. Thirdly, how to acquire measurements or image features automatically and robustly from frame images sequence is still a problem to be further researched.

2 Related work

2.1 Initialization

Initialization of 3D freehand pose is always related to the problem of single-frame pose estimation. A person-independent recognition method for hand postures against complex backgrounds is proposed in [5] by combining different feature types at the graph nodes. To estimate arbitrary 3D freehand postures, N. Shimada [6] accepts not only predetermined hand signs but also arbitrary postures in a monocular camera environment. The estimation is based on a 2D image retrieval. More than 16,000 possible hand appearances are originated from a given 3D shape model by rotating model joints and storing them in an appearance database. A Specialized Mappings Architecture (SMA) approach, proposed by Rómer Rosales [7], is to map image features to likely 3D hand poses by employing a machine learning architecture, with the help of rotation and scale-invariant moments of the hand silhouette. The SMA's fundamental components are a set of specialized forward mapping functions and a single feedback matching function. The joint angle data in the training set is obtained via a CyberGlove, a glove with 22 sensors that monitor the angular motions of the palm and fingers. During training, the visual features are generated by using a computer graphics module that renders the hand from arbitrary viewpoints given the 22 joint angles. Generally speaking, SMA provides continuous pose estimates through regression. The chamfer distance, edge orientation histogram and moment are used in [8] to estimate 3D hand shape and orientation by retrieving appearance-based matches from a large database of synthetic views, which are rendered by 26 predefined prototype shapes. The hand shape in the input image is assumed to be close to one of the 26 predefined shapes. A tree-based representation [9], where the leaves define a partition of the state space with piecewise constant density, can be applied effectively to track 3D articulated and non-rigid motion.

The single-frame pose estimation approach [10], based on a local search, expects to work well at the initial phase, because no more history information can be used. One of the distinct features of single-frame pose estimation approach is to retrieve hand poses from hand image database. This approach benefits from the appearance-based matching for 3D parameter estimation, by looking up the ground truth labels of the retrieved synthetic views. In [11], Stochastic Meta-Descent (SMD) algorithm is employed in an eight-particle tracker to track high-dimensional articulated structures with far fewer samples than in the previous methods, as well as to handle multiple hypotheses. Martinde LaGorce et al. [12] assumed that hand is parallel to the image plane at initialization and linear constraints are defined on relative length of the parts within each finger.

2.2 Cognitive behavioral model

Cognitive psychology is an approach in psychology that emphasizes internal mental processes. Cognitive models have been successfully used in three main ways by human–computer interaction (HCI) [13–15]. Cognitive models are used to modify interaction to help operators with their tasks [16, 17]. The cognitive behavioral model is based on the cognitions influence on behavior and vice versa, it is useful to HCI by predicting task times, assisting operators, and acting as surrogate operators. If cognitive models could interact with the same interfaces that operators do, the models would be easier to develop and apply as interface testers.

2.3 Visualization

Visualization [18] has been around for a long time. It is a process of transforming information into a visual form enabling the viewer to observe, browse, make sense, and understand the information. Through visual imagery, it has been an effective way to communicate both abstract and concrete ideas, and typically it employs computers to process the information and computer screens to view it using methods of interactive graphics and imaging [19].

2.4 HCI

As computers become integrated into everyday objects, effective natural human–computer interaction becomes critical: in many applications, operators need to be able to interact naturally with computers the way face-to-face human–human interaction takes place [20]. HCI design should consider many aspects of human behaviors [21]. The operator activity has three different levels: physical [22], cognitive [23], and affective [24]. The physical aspect determines the mechanics of interaction between human and computer while the cognitive aspect deals with ways in which operators can understand the system and interact with it. The affective aspect is a more recent issue and it lies not only in making the interaction a pleasurable experience for the operator but also in affecting the operator in a way that makes him or her continue to use the machine by changing attitudes and emotions towards the operator [25]. The ultimate goal of HCI is to provide an easier, more pleasurable and satisfying experience for the operator. By now, we have not found a way such as ours, which blends techniques of computer interaction, visualization and theories of cognitive behavioral models, to settle down initialization issue for 3D hand-tracking models.

Summarily, the above approaches, no matter how 3D tracking or 3D reconstruction algorithm works, are not so flexible as to satisfy the need of novel HCI in VR system. For example, many hand-tracking systems, based on image sequences or off-line videos, require the initial poses and positions identified manually to be the same as those in a hand pose data set.

The main contribution of our work consists in fusing the three core techniques, human–computer interaction in initializing, 3D hand model visualization and the operator's cognitive behavioral modeling, in an attempt to improve accuracy and reduce computational cost derived from the high-dimensional articulated hand structure.

3 The proposed method

3.1 Problem definition

We attempt to find a solution of V and R for the following problem:

$$\begin{cases} \underset{V,R}{\operatorname{Min}} E\{V, R\} \\ V_0, R_0 \\ E(V, V_0) < \Phi \end{cases} \tag{1}$$

where R is an operator's hand pose from an online video and V is a synthesized 3D virtual hand model which is synthesized by the computer based on freehand model, V_0 and R_0 are the initial synthesized 3D virtual hand model and real hand pose, respectively. In the expression (1), $E(\cdot)$ is the cost function used to evaluate the similarity of V and R, $\operatorname{Min} E(\cdot)$ is the 3D virtual hand model which is nearly the same as the operator's real hand pose, including both their positions and poses, Φ is a threshold value.

3.2 Overview

Just as defined in the formula (1), the objective of 3D-freehand-pose initialization is to reconstruct 3D hand models in accordance with the operators' poses based on their hand images from calibrated monocular cameras.

Our method is composed of pose recognition, coarse-tuning for hand poses and fine-tuning for 3D models (see Fig. 1).

The objective in hand pose recognition is to retrieve a rough 3D hand model, which is similar to the operator's initial hand pose. The objective of coarse-tuning process is to move the operator's hand towards the projection of the 3D hand model while it keeps fixed, and the objective of fine-tuning process is to fine-tune the 3D hand model making it be the same as the operator's hand while the operator's hand keeps fixed.

In order to provide a new approach to initializing 3D freehand model which is human-oriented, the following three core techniques are for the first time being blended together in our method: human–computer interaction in the process of initializing, visualization of 3D hand model and modeling the operator's cognitive behaviors.

The key of this paper is as follows. First of all, by means of blending cognitive psychology and case observations, cognitive behavioral models for operators are set up. Then, a novel 3D-freehand-pose initialization algorithm based on operator's cognitive behavioral models is proposed. Lastly, experiments are performed to demonstrate the performance of our method.

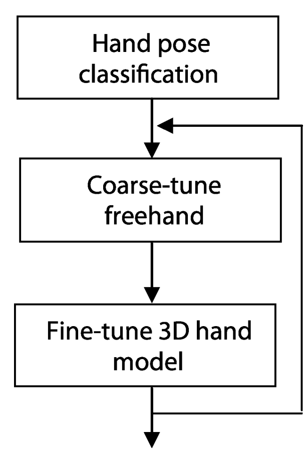


Fig. 1 The proposed 3D model initialization approach. Our algorithm is featured with interaction between the bare hand operator and the computer in the process of initialization

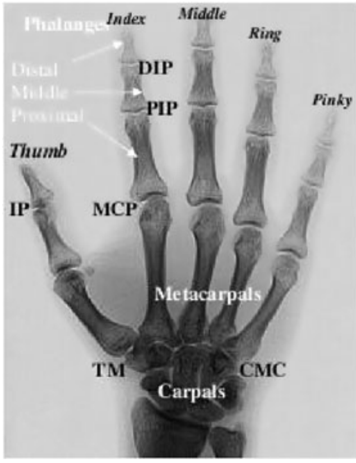


Fig. 2 The kinematic hand model [1]

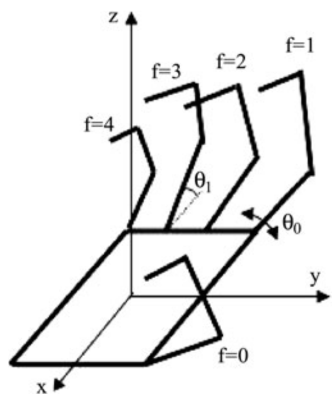


Fig. 3 The proposed 3D hand model with 26 encoded DOF

3.3 Hand mode

Based on the kinematic model [1] shown in Fig. 2, the 3D freehand model designed for initialization is shown in Fig. 3. The freehand consists of 27 bones, 8 of which are located in the wrist. The other 19 constitute the palm and fingers. The bones in the skeleton form a system of rigid bodies connected together by joints with one or more degrees of freedom per rotation. Each knuckle is simulated by a column with fixed radius and length, and the palm is simulated by a cuboid. In our model, all joints are located in the same plane on a finger, the plane is perpendicular to the palm plane with thumb excluded. Any two fingers are not allowed to overlap. There are 26 degrees of freedom (DOF): 20 local DOF and 6 global DOF. The fingers and the thumb have two DOF at the anchoring joint allowing flexion with an angle as well as spreading movement. The proximal interphalangeal and the distal interphalangeal joints have one DOF each. The kinematic constraints are listed in Table 1.

Table 1 The kinematic constraints of hand joints (the unit: degree)

f	θ			
	$\theta_0^{\min}$	$\theta_0^{\max n}$	$\theta_1^{\min}$	$\theta_0^{\max}$
0	−30	20	−20	40
1	−10	10	0	90
2	−10	10	0	90
3	−10	10	0	90
4	−10	10	0	90

3.4 Modeling cognitive psychology

How should we make the initialization process pleasurable and amusing? Researching the behavioral models of operators' hands is an important way to achieve this purpose.

According to cognitive psychology [14], the behavior of freehand is controlled by cognitive model, which reflects the cognitive processes and disciplines of human psychology, and this cognitive model keeps stable in the period of completing a specific cognitive task. A human performs cognitive tasks according to mental models.

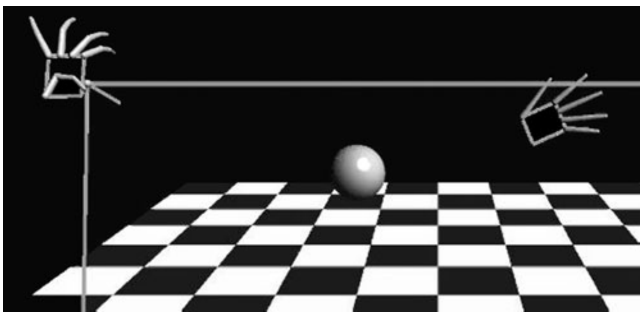
GOMS (Goals, Operations, Methods, and Selection rules) models [26] are a well-developed tools for mental evaluation in their current modern form, and we find that it is still applicable for describing cognitive psychology models in our initialization system. We combine GOMS and theory of cognitive load to model cognitive behaviors, placing emphasis upon the way to describe Operations and Methods in GOMS by the means of observation and case analysis.

3.5 Case investigation

In order to study Operations and Methods in GOMS and acquire steady Cognitive Behavioral Model (CBM) in the process of hand pose initialization, the participators were invited to take part in our experiments: each of them was equipped with a data glove on his/her right hand. A 3D hand model was visualized in the scene and each participator was requested to adjust his/her right hand from an initial natural pose until the hand pose and the 3D model were the same. In this process, all data from data glove sensors is reserved for further analysis. A research scene on CBM is shown in Fig. 4.

We just present the angle curves with time for the middle finger and the forward direction vector of the palm shown in Fig. 5: they are just a little clip of the 26 curves used in analysis of CBM. We observed from Fig. 5 that the curves can be divided into two parts, the one acutely changing in the first period of time, which can be assigned to a coarse-tuning process; and the other placidly changing in the second period of time, which can be assigned to a fine-tuning process. This observation stimulates us to introduce the technique of

(a)



(b)

Fig. 4 A research scene on CBM. (**a**) The operator was equipped with a data glove on his right hand. (**b**) The hand on the right is the fixed 3D hand model and on the left is the synthesized 3D virtual hand based on data glove data. We attempt to probe into some clues about how operators change their hands for superposition of the synthesized hand and the 3D hand model

interaction between human and computer, and it also provides us with cues to assign interactive and cooperative tasks to the computer and the operator based on the two periods. The CBM models are stated in the form of CBM features.

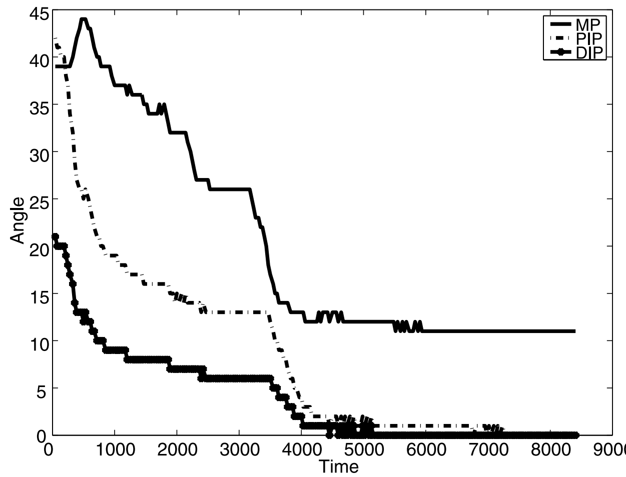
3.6 CBM features

Observation and analysis of the series of experiments yields the following CBM featured points:

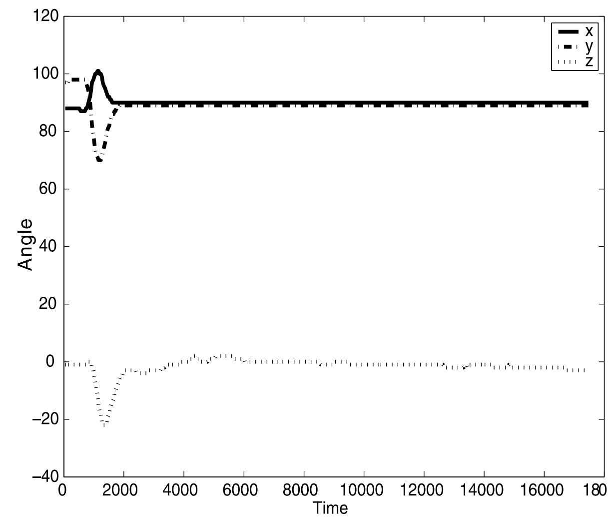
CBM-1: In the process of adjusting variables of pose vectors, changes with large range occur in frontal period of time while changes with small range occur in back period of time.

CBM-2: A hand pose vector is always composed of variable parts and unvariable parts. Taken gun-style pose as an example, in the whole process of initialization, the middle finger, the ring finger and the little finger are kept unvariable.

CBM-3: Compared with the approach of asking operators to imagine and form their required hands only according to required regulations or constraints, allowing the operators to actively adjust their hand poses towards visible 3D hand models shown on displays is more coincident with operators' cognitive customs and less with their cognitive loads.



(a)



(b)

Fig. 5 (**a**) The change curves with time for a middle of the finger. (**b**) The forward direction vector of the palm expressed as the angles of forward direction vector of the palm with x-, y- and z-axes. The curves start with the initial hand gesture and end with the situation when freehand returns to the initial state. The operator was required to wear a data glove with hand-position tracker system on his/her right hand. The time unit is millisecond. The angle unit is degree

CBM features are based on interaction between operators and computer.

3.7 Fine-tuning process

The superiority of operator over the computer is that he/she can flexibly make decisions to all the situations in the process of completing a cognitive task, so it is reasonable to introduce the technique of interaction between human and computer into our initialization system. How to make computers to adapt to humans, or how to make initialization process easier, more pleasurable in experience for the operators, is the main objective in our system.

Operator's initial freehand pose is always far from the predefined one, meaning that to search the solution for the problem definition (1) requires computationally expensive searching algorithm. In order to deal with this problem, we assign the operator to do this intelligent work corresponding to the frontal period of time stated in CBM-1. The operator adjusts his/her bare hand's position and pose in agreement with CBM-2 to make the online freehand images to be superposed onto the projection of the given synthesized initial 3D hand model while the 3D hand model is kept fixed.

Once the operator's freehand fall into the neighborhood of the 3D freehand model, the computer will provide the operator with feedback, flickering and highlighting, and the computer begins to fine-tune the 3D hand model using the approach similar to a particle filtering (PF) [27] in which the CBM-2 is followed in order to greatly reduce the particle number. Fine-tuning is assigned to the back period of time stated in CBM-1.

The above two phases go by turns between human and computer until the operators feel satisfied or the Hausdorff distance between the human image and the projection of the 3D hand model onto the image is in the required scope.

According to CBM-3, the human–computer interaction to some extent depends on the way of feedback and effectiveness of visualization. In our study, many approaches are used for visualization. For example, the 3D hand model is rendered and outputted by means of OpenGL, the new hand images are displayed in real time with a visual style.

What the computer does is to adjust 3D hand model once the operator accomplishes adjusting his/her freehand. Even if approximate poses are obtained by pose classification, it is still difficult for the computer to further fine-tune 3D freehand models with fast speed until the 3D hand model is the same as the operator's hand pose because of high-dimensionality problem. It is fortunate that CBM-2 can help us alleviate this problem. For example, if the computer knows or predicts that some part of the variables in a hand pose vector will change with time, it will focus on identifying the values of these variables and pay little attention to the left variables to be unchanged with time. This approach is equivalent to reducing dimensionality of the hand pose vector.

The fine-tuning process by computer is as follows:

(1) Generate N particles X_1^i of X_1, $i = 1, 2, \ldots, N$, using Gaussian model in agreement with the CBM features.

(2) Compute weight ω of each particle by

$$h^{(i)} = \text{Hausdorff}\bigl(X_1^{(i)}, \Omega\bigr), \tag{2}$$

$$s\_h^{(i)} = e^{-h^{(i)}}, \tag{3}$$

$$\omega_i = \frac{s\_h^{(i)}}{\sum_{j=1}^{N} s\_h^{(i)}}. \tag{4}$$

(3) State updating

$$X = \sum_{i=1}^{N} \omega_i X_1^{(i)}. \tag{5}$$

(4) X is evaluated by

$$E = \text{Hausdorff}(X, \Omega). \tag{6}$$

In the formulas (2) and (6), Hausdorff(X, Ω) is the Hausdorff distance between the projection of hand model X onto hand image plane and the hand image. The hand image features are extracted by multiple scale approach [28].

4 Experimental results

4.1 Experimental settings

We use a color CCD calibrated camera ZT-QCO12 with a 4 mm lens that captures 640 × 480 video at 30 Hz. Our computer is Intel®, Core™, Quad CPU 2.66 GHz, 3.25G memory. We employ a 3D hand model with 26 degrees of freedom (DOF), 6 DOF for the global transformation and 4 DOF for each finger. The length of each knuckle on the finger and the size of the palm are determined beforehand.

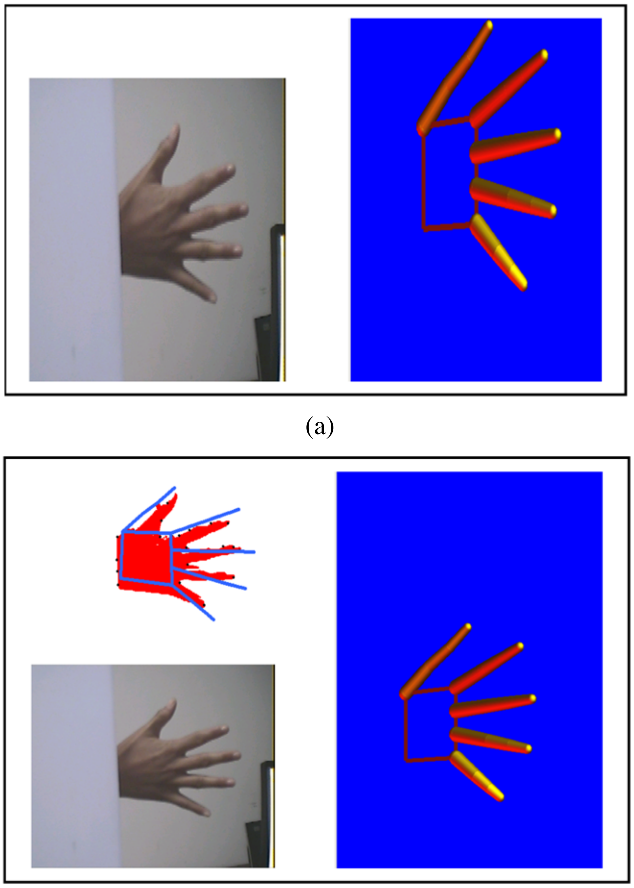
4.2 Initialization of 3D hand model

Our method is compared with the two widely used approaches, the single-frame pose estimation approach [10] and the SMD approach [11], and for simplification, SF and OM stand for the single-frame pose estimation approach and our proposed method, respectively.

The 3D-freehand-pose initialization process by OM is shown in Fig. 6.

This process is composed of the three phases: hand pose recognition [29], coarse-tuning and fine-tuning. The coarse-tuning process is performed by the operator and the fine-tuning process is performed by computer.

We used three quantitative measures to evaluate initialization algorithms: accuracy, time cost and cognitive burden. Accuracy is evaluated by Hausdorff distance between hand images and projection of 3D hand models onto the image planes (which is also called as Hausdorff distance in the remaining part of this paper for simplicity). Cognitive burden was evaluated by the four factors: tiredness, joviality, freedom and workability. The tiredness is described the extent of toil the user experiences in the process of initialization; the joviality describes the degree of amusement the user feels; the convenience describes the quality of being suitable to user's purposes; and workability describes the extent to which the initialization approach is feasible. The four factors are scored between 0 and 100 by the users.



(a)

(b)

Fig. 6 The 3D-freehand-pose initialization process by OM. (**a**) Pose recognition [29], obtaining an initial 3D model. (**b**) The process of interaction between human and computer. In (**b**), *the top-left* is the visualization of 3D hand model and image features. *On the right* is the initialized 3D hand model. *The bottom-left* is a frame image from the monocular camera

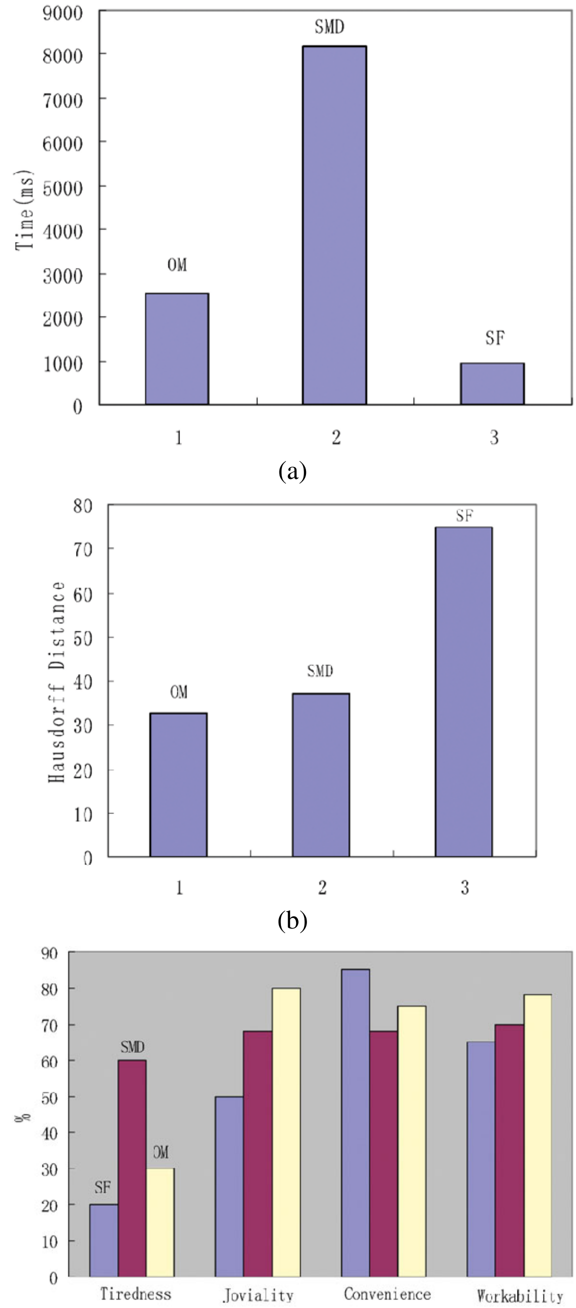
We asked an operator to accomplish an initialization task, using OM, SF and SMD respectively, 50 times for each, and the average results are shown in Fig. 7. The time costs by OM, SMD and SF are 2535.3 ms, 8176.6 ms and 936.66 ms respectively (Fig. 7a). Hausdorff distances of OM, SMD and SF are 32.569, 37.205 and 74.825 respectively (Fig. 7b). OM has the least Hausdorff distance of the three methods.

The investigation of cognitive burden of OM, SF and SMD was conducted and the results are shown in Fig. 7c. It is shown that OM is easiest, most pleasurable for the operators for the online use because of the least cognitive burden.

The Hausdorff distances in each of 50 times are shown in Fig. 8. On a whole, the Hausdorff distance of OM is the least.

We further carried experiments in the case of different operators. Fifty students were asked to perform the same experiment, the experimental results are shown in Fig. 9.

Figure 9 demonstrates that the evaluation results are according to the same operator.



(a)

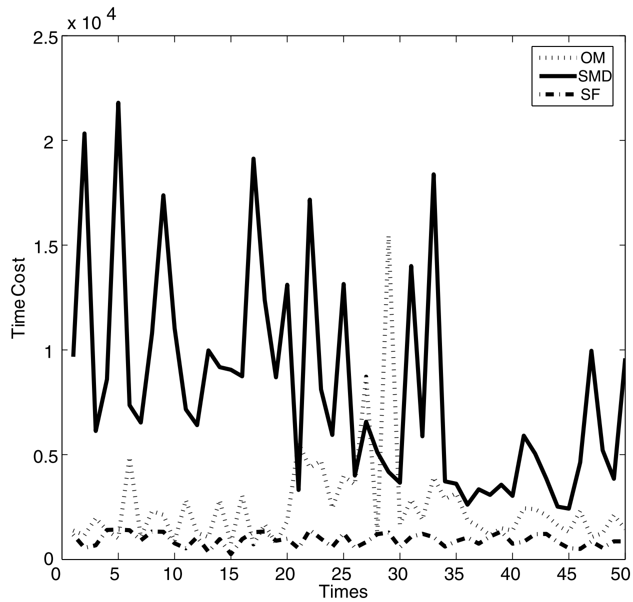
(b)

(c)

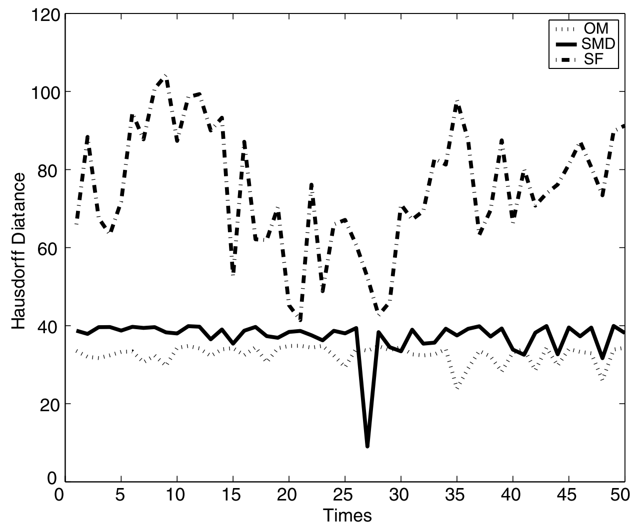
Fig. 7 The average time cost, accuracy and feedback for cognitive load, with the same operator participating in the experiments OM, SF and SMD, 50 times in each. (**a**) The average time of OM, SF and SMD. (**b**) The average accuracy of OM, SF and SMD. Accuracy is evaluated by Hausdorff distance. (c) The average tiredness, joviality, freedom and convenience of OM, SF and SMD. These experiments were performed by the same operator

4.3 Analysis of the experimental results

Our initialization system benefits much from visualization. Because of the use of visualization, it is possible that human–computer interaction is carried through harmoniously, and that the operators can feel intuitively how the



(a)

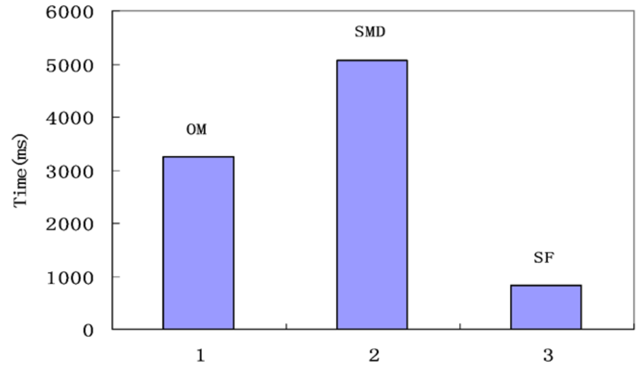


(b)

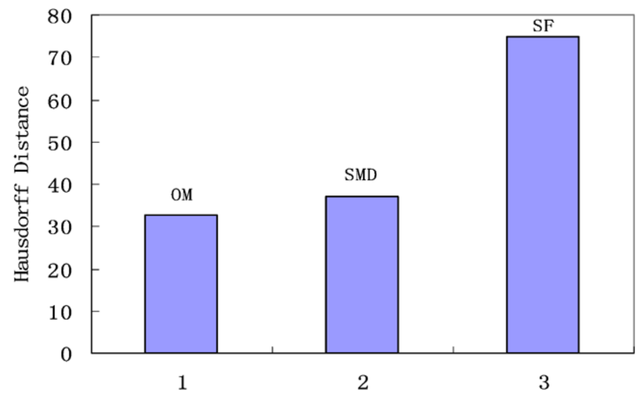
Fig. 8 The experimental results of 50 times by the same operator. (**a**) The time cost of OM, SF and SMD. (**b**) The Hausdorff distance in each of 50 times for OM, SF and SMD. Hausdorff distance is used to evaluate 3D hand models after initialization is performed. These experiments were done by the same operator

orientations and ranges of their hands should be adjusted, no matter how big the distance between the objective position and the current position is.

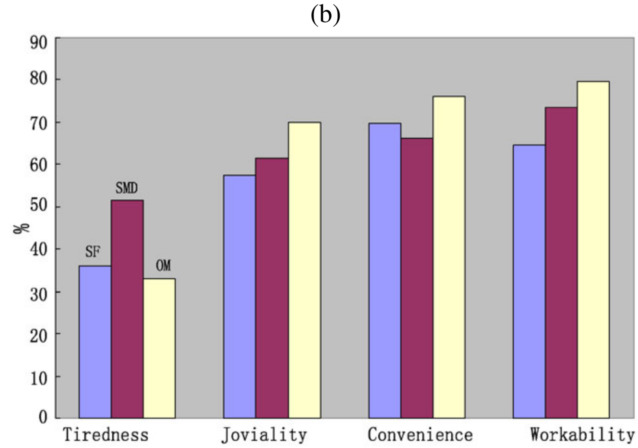
The time cost of our algorithm is composed of the three parts: the time cost in hand pose recognition by computer, in coarse-tuning process by operators and in fine-tuning process by computer. The time cost by coarse-tuning process depends mainly on skills for the operator to manipulate our initialization system, while the time cost by computer is mostly impacted on the number of particles used because



(a)



(b)



(c)

Fig. 9 For 50 different operators, the average curves of time cost, accuracy and feedback for cognitive load. (**a**) The average time of OM, SF and SMD. (**b**) The average accuracy of OM, SF and SMD. (**c**) The average tiredness, joviality, freedom and convenience of OM, SF and SMD

of high dimensionality of a 3D hand structure. On the other hand, based on the CBM-2, the dimensionality of freehand pose vectors is greatly reduced.

What we benefit most from CBM and interactive behavior between human and computer is that it makes our initialization system fast, accurate, robust and intelligent, as well as human-oriented for operating, convenient for online use, beneficial to reducing cognitive loads of operators. These

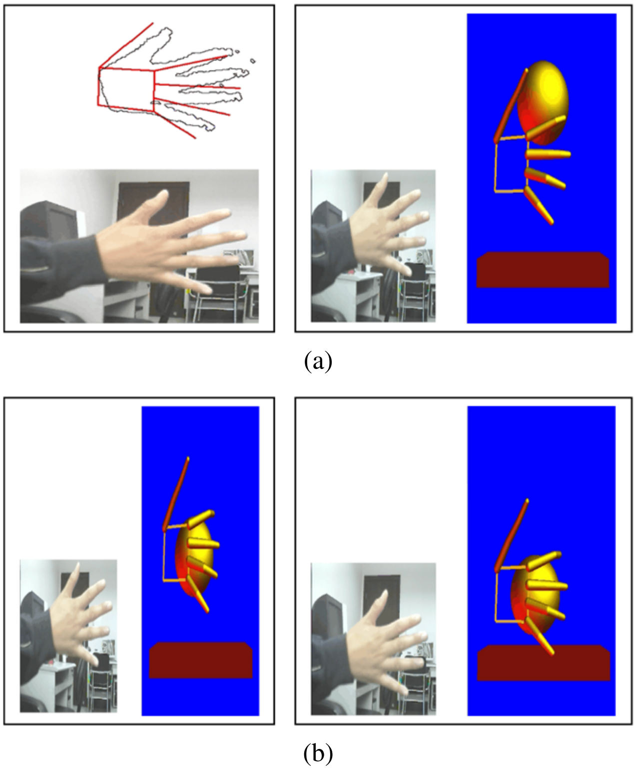


Fig. 10 The operator is moving a vehicle part from one place to another with his bare hand. The first scene is in the process of initialization of freehand. The last three scenes are tracking process by PF [27], our operator are moving a vehicle part from one place to another in our assembly system. This experiment is performed in complex background

superiorities over other initialization systems are important for HCII (Human Computer Intelligent Interaction) or PUI (Perceptual User Interface), and are also parts of the objectives in our assembly system.

4.4 Application of OM

Our initialization system has been applied to our bare 3D freehand tracking which is a part of our virtual assembly system (see Fig. 10).

5 Discussion and conclusion

Human–computer interface with intelligence, nature, amusingness and convenience is one of the important components in our virtual assembly system. Tracking, recognition and interaction based on bare 3D freehand are a part of the cores in our assembly system. The freehand pose tracker requires a reliable initial pose in the first frame. Estimating the freehand pose in a single frame without a strong prior of the freehand pose is challenging. Unfortunately, initialization has not been paid enough attention for many years and there are few related or specialized papers or products available to retrieve from or to refer to.

One motivation for addressing this challenging problem is for the purpose of initializing tracking without imposing too many constraints on the user and making initialization human-oriented for operating, convenient for online use, and beneficial to reducing cognitive loads of operators.

A novel algorithm for initializing human 3D freehand model is proposed in this paper. The main contribution of this paper is that the three techniques—interaction between human and computer, modeling cognitive behaviors for operators, and visualizing information in the process of initialization—are fused together to meet the needs of fast speed, high accuracy, high intelligence.

The visualization system based on computer vision with monocular camera has been implemented with VC++6.0, and many experimental results demonstrated that our system is direct, amusing, natural, intelligent and convenient. Furthermore, this system has successfully been applied to our 3D freehand tracking system and simulation of virtual assembly system.

Cognitive behavior in itself is very complex. How to intensively explain, analyze and model cognitive behavior is intended as our future work.

Acknowledgements This paper is supported by NSFC (No. 60773109), NSFC (No. 60973093), Natural Science Foundation of Shandong Province (Y2007G39), Natural Science Foundation for Distinguished Youth Scholar of Shandong Province (No. JQ200820), Key Project of Natural Science Foundation of Shandong Province (2006G03), Science and Technology Plan of Shandong Province Education Department (J07YJ18).

References

1. Erol, A., et al.: Vision-based hand pose estimation: a review. Comput. Vis. Image Underst. **108**, 52–73 (2007)
2. Julier, S.J., Uhlmann, J.K.: A new extension of the Kalman filter to nonlinear systems. In: Procedure of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, pp. 82–193. SPIE, Bellingham (1997)
3. Cui, J.S.: Studies on three-dimensional model based posture estimation and tracking of articulated objects. PhD Thesis, Tsinghua University, Beijing, China (2004)
4. Erol, A., Bebis, G., Nicolescu, M., Boyle, R., Twombly, X.: A review on vision-based full DOF hand motion estimation. In: Proceedings of the IEEE Workshop on Vision for Human–Computer Interaction (V4HCI), San Diego, California, vol. 3, pp. 75–83 (2005)
5. Triesch, J., von der Malsurg, C.: A system for person-independent hand posture recognition against complex background. IEEE Trans. Pattern Anal. Mach. Intell. **23**(12), 84–95 (2001)
6. Shimada, N., Kimura, K., Shirai, Y.: Real-time 3-D hand posture estimation based on 2-D appearance retrieval using monocular camera. In: Proc. Int. Workshop RATFG-RTS. pp. 23–30 (2001)
7. Rosales, R.: The specialized mappings architecture with applications to vision-based estimation of articulated body pose. PhD Thesis. Boston University, Boston (2002)

8. Athitsos, V., Sclaroff, S.: Estimating 3D hand pose from a cluttered image. Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn. **2**, 432–439 (2003)
9. Stenger, B., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Filtering using a tree-based estimator. Proc. IEEE Int. Conf. Comput. Vis. **2**, 1063–1070 (2003)
10. Tomasi, C., Petrov, S., Sastry, A.: 3D tracking = classification + interpolation. In: The Ninth IEEE International Conference on Computer Vision, pp. 1441–1448 (2003)
11. Bray, M., Koller-Meier, E., Goo, L.V.: Smart particle filtering for 3D hand tracking. In: The Sixth IEEE International Conference on Automatic Face and Pose Recognition, pp. 675–680. IEEE Computer Society, Los Alamitos (2004)
12. LaGorce, M., Paragios, N., Fleet, D.: Model based hand tracking with texture, shading and self-occlusions. IEEE Proc. IEEE Conf. Comput. Vis. Pattern Recogn. pp. 1–8 (2008)
13. John, B.E.: Cognitive modeling in human–computer interaction. In: Proceedings of Graphics Interface, pp. 161–167 (1998)
14. Card, S.K., Moran, T.P., Ewell, A.: The Psychology of Human–Computer Interaction. Lawrence Erlbaum Associates Inc., Hillsdale (1993)
15. John, B.E., Kieras, D.E.: Using GOMS for operator interface design and evaluation: which technique? ACM Trans. Comput. Hum. Interact. **3**(4), 287–319 (1996)
16. Anderson, J.R., et al.: Cognitive tutors: lessons learned. J. Learn. Sci. **4**(2), 167–207 (1995)
17. Ritter, F.E., Baxter, G.D., Jones, G., Young, R.M.: Supporting cognitive models as operators. ACM Trans. Hum. Comput. Int. **7**(2), 141–173 (2000)
18. Tang, Z.: Visualization of 3D Datasets [M]. Singhua University Press, Beijing (1999)
19. Cleveland, W.S.: Visualizing data [M]. Hobart Press, New Jersey (1993)
20. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: a survey. Comput. Vis. Image Underst. **108**(1–2), 116–134 (2007)
21. Karray, F., Alemzadeh, M., Saleh, J.A., Arab, M.N.: Human–computer interaction: overview on state of the art. Int. J. Smart Intell. Syst. **1**(1), 137–159 (2008)
22. Chapanis, A.: Man Machine Engineering. Wadsworth, Belmont (1965)
23. Norman, D.: Cognitive engineering. In: Norman, D., Draper, S. (eds.) Operator Centered Design: New Perspective on Human–Computer Interaction. Lawrence Erlbaum, Hillsdale (1986)
24. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
25. Te'eni, D., Carey, J., Zhang, P.: Human Computer Interaction: Developing Effective Organizational Information Systems. Wiley, Hoboken (2007)
26. John, B.E., Kieras, D.E.: The GOMS family of operator interface analysis techniques: comparison and contrast. ACM Trans. Comput. Hum. Interact., 320–351 (1996)
27. Gordon, N., Salmond, D.J., Smith, A.F.M.: Novel approach to nonlinear and non-Gaussian Bayesian state estimation. IEE Proc. F **140**, 107–113 (1993)
28. Feng el al, Z.: Research on features extraction from frame image sequences. In: International Symposium on Computer Science and Computational Technology (ISCSCT'2008), pp. 762–766 (2008)
29. Stefan, A., Athitsos, V., Alon, J., Sclaroff, S.: Translation and scale-invariant gesture recognition in complex scenes. In: Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments, Athens, Greece, pp. 1–8 (2008)

Zhiquan Feng is a professor of School of Information Science and Engineering, Jinan University. He received the Master's degree from Northwestern Polytechnical University, China in 1995, and PhD degree from Computer Science and Engineering Department, Shandong University in 2006.
He has published more than 50 papers on international journals, national journals, and conferences in recent years. His research interests include: human hand tracking/recognition/interaction, virtual reality, human–computer interaction and image processing.

Minming Zhang is an Associate Professor Computer and Engineering Department, Zhejiang University. She got the Bachelor's degree from Computer Science Department, Nanjing University in 1990, and the Master's and PhD degrees from Computer Science and Engineering Department, Zhejiang University in 1995 and 2008.
She has published more than 30 papers on international journals, national journals, and conferences in recent years.
She is the co-author of two books related to computer graphics and multimedia. Her research interests include: virtual reality/virtual environment, multi-resolution modeling, real-time rendering, distributed VR, visualization, multimedia and image processing.

Zhigeng Pan received his Bachelor's and Master's degrees from the Computer Science Department in 1987 and Nanjing University, and PhD degree from Zhejiang University in 1993.
Since 1996, he has been working at the State Key Lab of CAD&CG as a Full Professor. He is a member of SIGGRAPH, Eurographics, IEEE, a senior member of the China Image and Graphics Association. He is on the director board of the International Society of VSMM (Virtual System and Multimedia), a member of IFIP Technical Committee on Entertainment Computing (acting as representative from China).
Currently, he is the Editor-in-Chief of the International Journal of Virtual Reality. He is on the editorial board of International Journal of Image and Graphics, International Journal of CAD/CAM, Journal of Image and Graphics, Journal of CAD/CG. He is the program co-chair of EGMM'2004 (Eurographics workshop on Multimedia), Edutainment'2005 and VEonPC'2005, and is the program co-chair of Edutainment'2006, conference co-chair of ICAT'2006, Cyberworlds'2008.

Bo Yang is a Professor and Vice-president of University of Jinan, Jinan, China. He is the Director of the Provincial Key Laboratory for Network-based Intelligent Computing and also acts as the Associate Director of Shandong Computer Federation, and Member of the Technical Committee of Intelligent Control of Chinese Association of Automation. His main research interests include computer networks, artificial intelligence, machine learning, knowledge discovery, and data mining. He has published numerous papers and gotten some of important scientific awards in this area.

Tao Xu was born in 1979, is currently a lecturer in the School of Information Science and Engineering, University of Jinan, Jinan, Shandong, PRC. His research interests include computer vision, object tracking, pattern recognition. He received his BSc degree in School of Information Science and Engineering from University of Jinan in 2001, his MSc degree in College of Computer and Communication Engineering from China University of Petroleum, Dongying, Shandong, in 2007.

Haokui Tang was born in 1972, his research interests include image analysis and understanding, computer vision.

Yi Li female, born in 1961, Associate Professor, Master Supervisor. Her main research interest is information control.