Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

SSC: Gesture-based game for initial dementia examination*

LIU Jun-fa[†1,2], CHEN Yi-qiang[1], XIE Chen[3], GAO Wen[1,2]
([1]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China*)
([2]*Graduate School of the Chinese Academy of Sciences, Beijing 100080, China*)
([3]*College of Software Technology, Zhejiang University, Hangzhou 310027, China*)
[†]E-mail: jfliu@jdl.ac.cn

Received Apr. 3, 2006; revision accepted Apr. 17, 2006

Abstract: This paper presents a novel system assisting medical dementia examination in a joyful way: the object just needs to play a popular game SSC against the computer during the examination. The SSC game's target is to detect the player's reacting capability, which is related closely with dementia. Our system reaches this target with some advantages: there are no temporal and spatial constraints at all. There is no cost, and it can even improve people's mental status. Hand talk technology and EHMM gesture recognition approach are employed to realize the human computer interface. Experiments showed that this system can evaluate people's reacting capability effectively and is helpful for initial dementia examination.

Key words: Human computer interface, Hand talk, Dementia, Reaction capability
doi:10.1631/jzus.2006.A1253 **Document code:** A **CLC number:** TP391

INTRODUCTION

Dementia is becoming a big problem in modern life, and particularly, it takes place more frequently in mental job group and the elderly. Today, more than 20 million people in the world catch this disease (Francis *et al*., 1999; Almkvist *et al*., 1998). So how to detect underlying dementia and take preventive measure is critical for both the doctors and their patients.

People with dementia always behave slowly and their memory decreases with the development of the disease. Currently, the traditional measure to examine the dementia of the patient is to undergo WAIS-R (Wechsler Adult Intelligence Scale-Revised), WAMS-R (Wechsler Adult Memory Scale-Revised), or HDS (Hasegawa Dementia Scale) test (Guo *et al*., 2000; Geldmacher and Whitehouse, 1996; Kawas, 2003; Knopman *et al*., 2001). Such treatments are dedicated and exact for measuring people's memory or intelligence status. However, there are still some practical problems worthy of note. First, it costs much for the expensive instrument and comprehensive procedure, and second, some of the parameters are sensitive only to seriously ill Alzheimer's patient. Furthermore, the procedures are time-consuming (lasts more than 10 min), and affected by the object's specific reaction when receiving the examining, and so on.

This paper proposes a game system to help dementia detection. Different from other medical measurements, this system concentrates on people's reacting capability. As we know, people with dementia move slowly or behave erratically as their brain cells are damaged in local place. So we can get some information on a disease from people's reacting capability.

The game SSC (Scissors, Stone, Cloth) is a very popular program being adopted when people want to become the winner in daily life. In our system, one of the competitors is the computer; the other is the people who may want to investigate his/her own reacting capability. For examination, our game system has some advantages: there is no cost for the player, and

* Project supported by the National Nature Science Foundation of China (Nos. 60303018 and 60533030) and Beijing Science and Technology New Star Project (No. 2005B54), China

the player can do the examination at home by himself whenever he wants to, and the procedure is short and joyous, so the player does not have to behave in any specific way. Finally, the game is so popular that almost everyone had experienced it. The player can make a primary brain health evaluation, then decide if he needs to take a professional medical test.

As shown in Fig.1, the system comprises three modules, which are hand gesture synthesis module, hand gesture recognition module and scoring module. The first module displays a virtual image on the screen and generates three hand gestures representing Scissors, Stone, and Cloth, respectively. The second module captures the player's hand image via a camera and recognizes the gesture type. The last module makes a game program, records the game process, and rewards a score to the player.

Fig.1 The human computer interface

The following parts are organized as such: Section 2 explains how we synthesize the virtual player and its hand gestures. The recognition of the hand gesture will be explained in Section 3. In Section 4, we give the designing of the scoring module and the rules to evaluate the player. We will show the performance of the system for reacting capability detection through several groups of experiments in Section 5. In the last section, we draw conclusions on our system.

HAND GESTURE SYNTHESIS

Virtual human synthesis

Hand gesture synthesis consists of two aspect jobs. The first and precondition is the synthesis of the virtual human. We had realized the virtual environment and virtual human and applied the technology to sign language synthesis successfully (Yang *et al.*, 2003). Here, we build the virtual human model by 3D tools, such as 3D MAX, and output the model in VRML format. The VRML (Virtual Reality Modeling Language) is convenient for describing the virtual human and is compatible with various platforms. We select OPGL as the presenting tool to create virtual environment and synthesize virtual human by loading the VRML file. To prompt the friendly User Interface, the system provides five virtual human models. They are listed in Fig.2. People may choose one of them as their competitor.

Fig.2 Five virtual human models

Hand gesture synthesis

The traditional way to synthesize hand gesture is to capture the performer's motion data, and represent it by the virtual object (Chen *et al.*, 2002; 2003). In this project, we can just use the tool GetstureEdit to create the game gestures, "Scissors", "Stone", and "Cloth". GestureEdit is very easy to use software we developed to compose arbitrary hand gestures. Fig.3 shows 5 key frames generated with GestureEdit that can synthesize an animation of "Scissors". In com-

Fig.3 Five key frames of "Scissors" animation

petition, if the computer needs to give a gesture of "Scissors", it only needs to load these five key frames and interpolate linearly some transitional frames to a piece of animation.

Gesture data reusing

The gestures are edited first based on and suitable for a dedicated virtual human model. To reuse the data of the other model, the retarget problem need to be handled. In (Ge *et al*., 2005), we proposed a retarget approach to solve this problem. First many sensitive points are defined on the human body, and key sensitive points and secondary sensitive points are specified through analyzing the importance degree of those points. Then a novel method based on the relative position of the original sensitive points is adopted to compute the target sensitive point position. Finally an adaptive IK (inverse kinematics) solver is utilized to realize the retargeting problem. Fig.4 shows the effect of retargeting. Fig.4a shows the original gesture, which is edited well by GestureEdit. Fig.4b is the data reused in another model without retargeting and Fig.4c shows results after retargeting.

(a) (b) (c)

Fig.4 Gesture retargeting effect. (a) The original gesture; (b) The data reused in another model without retargeting; (c) Results after retargeting

HAND GESTURE RECOGNITION

EHMM recognition algorithm

In this paper, we take EHMM (Embedded Hidden Markov Model) as recognition algorithm, which had been applied to image pattern recognition problems successfully (Ara and Monson, 1999; Kuo and Agazzi, 1994). Compared to traditional HMM for recognition task (Rabiner, 1989; Zhang *et al*., 2004), its advantage is finding structure information on the image. At this point, the gestures of "Scissor", "Stone" and "Cloth" can be distinguished well due to their different shapes.

EHMM is a pseudo two dimension HMM. Each EHMM is composed of some super states which behave as the states of a macro HMM, and each super state contains a micro HMM. The EHMM structure is showed in Fig.5.

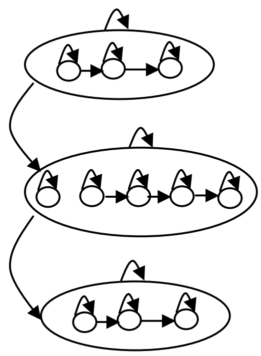

Fig.5 Structure of EHMM

Its form description is like this:

$$\lambda=(\pi, \boldsymbol{A}, \Lambda),$$

where:

(1) π is the initial probability of super states. $\pi=\{\pi_i, 1\le i\le N\}$, N is the number of super states.

(2) $\boldsymbol{A}$ is the state transition probability matrix. $\boldsymbol{A}=\{a_{ij}, 1\le i, j\le N\}$.

(3) Λ is a set of super states. $\Lambda=\{\Lambda^i, 1\le i\le N\}$, where:

$\Lambda^i=\{\pi^i, \boldsymbol{A}^i, B^i\}$ is a set of parameters that define the ith embedded HMM.

π^i is the initial probability of ith states embedded in HMM.

$\boldsymbol{A}^i=\{a_{kl}^i, l\le k, l\le N^i\}$ is the state transition probability matrix of ith embedded HMM. N^i is the state number of ith embedded HMM.

$B^i=\{b_k^i(\boldsymbol{O}_{t_0,t_1})\}$ is the sub-states' probability density function of ith EHMM, where $\boldsymbol{O}_{t_0,t_1}$ represents the observation vector at row t_0 and column t_1 (t_0=1, ..., T_0; t_1=1, ..., T_1). B^i is typically used as a finite mixture of the form

$$b_k^i(\boldsymbol{O}_{t_0,t_1})=\sum_{f=1}^{F} C_{kf}^i N(\boldsymbol{O}_{t_0,t_1}, \boldsymbol{u}_{kf}^i, \boldsymbol{U}_{kf}^i),\ 1\le k\le N^i,$$

where, F is the number of the mixture probability density function. C_{kf}^i is the mixture coefficient for the

fth mixture in kth embedded state and ith super state. $N(\boldsymbol{O}_{t_0,t_1}, \boldsymbol{u}_{kf}^i, \boldsymbol{U}_{kf}^i)$ is a Gaussian probability density function with mean vector $\boldsymbol{u}_{kf}^i$ and covariance matrix $\boldsymbol{U}_{kf}^i$.

Training and testing of EHMM

Once the Embedded HMM structure is defined, the next issue is to train it with the sample set. We collected a total of 300 pictures for each hand gesture as training samples. To adapt to the variance of the player's hand, the training samples are designed into three groups. Each group represents gestures with specialized angle. The gestures come from 70 males and 30 females.

Some samples are listed in Fig.6.

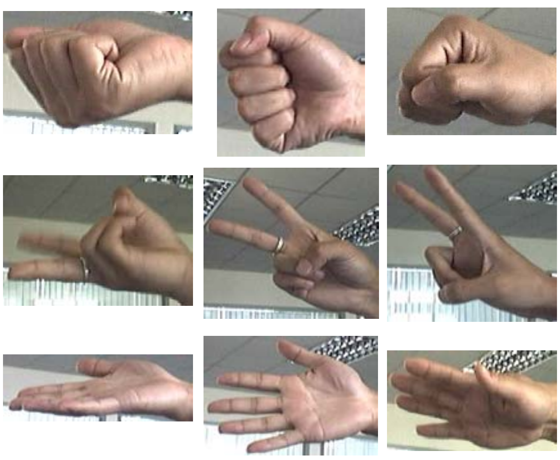

Fig.6 Gesture samples in three angles

We adopt DCT method to extract the frequency domain features, and then establish Gaussian Mixture Models with DCT coefficient. The training framework described in Fig.7, consists of the following steps:

(1) Initialize DCT parameters with mapping window size of 12×12 pixels, and step of 2×2 pixels to obtain DCT coefficient matrix.

(2) Assign uniform segmentation to DCT coefficient vectors to get initial segmentation. Calculate initial parameters for each block.

(3) Adopt doubly embedded Viterbi segmentation algorithm to generate new segmentation, and re-estimate the parameters using K-means algorithm.

(4) Redo Step (3) till the Viterbi segmentation likelihood of consecutive iterations is smaller than a threshold. The final EHMM is well trained.

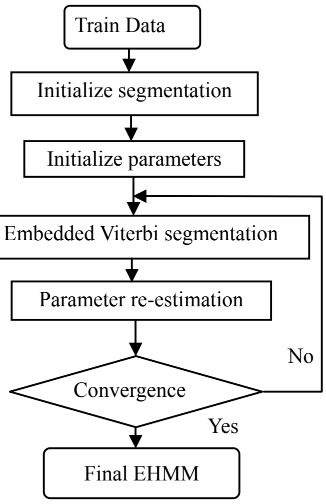

Fig.7 Training framework

In recognizing stage, when the unknown gesture image is given, we extract the DCT observation sequence vectors, and the probability of the observation sequence produced by an embedded HMM model is computed via a doubly embedded Viterbi recognizer. The model with the highest likelihood is selected and this model reveals the identity of the unknown gesture. When requiring the objects to play their hand gesture with some special constraints, we got recognition rate of 98.67% for the three gestures, which is high enough for actual application.

GAME SCORING

The SSC game system makes examination schedule in two modes. One mode is the learning examination; the other is the competing examination. In each mode, the hand gesture is synthesized and expressed in fast and slow speed. So there are a total of four phases in an experiment, which are SL (Slow Learning), FL (Fast Learning), SC (Slow Competing) and FC (Fast Competing). By default, in the phases of slow experiment, SL and SC, system synthesizes a hand gesture each second. In the phases of the fast experiment, FL and FC, the period for the system to synthesize a hand gesture is half second. Naturally, the time span can be adjusted. The four phases and their configurations are shown in Table 1.

Table 1 Four phases and their configurations

Item	Period	Score
SL (Slow Learning)	One second	1
FL (Fast Learning)	Half second	2
SC (Slow Competing)	One second	2
FC (Fast Competing)	Half second	3

We designed scoring rules for each case:

Learning Examination: The virtual player gives 10 gestures in random order, and the human player is asked to follow its gestures. Once giving a hand gesture, the system will capture an image of the player's hand in 500 ms, and then recognizes the hand gesture. If the two gestures are identical, the player may get one point in SL or two points in FL. In these two cases, one will get scores $Score_{\mathrm{SL}}$ and $Score_{\mathrm{FL}}$.

Competing Examination: In this mode, only when the player's hand gesture wins the virtual human's, can the player get points, and in SC case, the player gets two points, in FC case, gets 3 points. The rule to win is like this: Scissors wins Cloth, Cloth wins Stone, and Stone wins Scissors. That is a circle of power, but is an effective way to test one's reacting capability. In these two cases, $Score_{\mathrm{SC}}$ and $Score_{\mathrm{FC}}$ will be obtained.

Anyone who wants to undergo the examination is required to play the game and achieve total scores:

$$Score_{\mathrm{total}}=Score_{\mathrm{SL}}+Score_{\mathrm{FL}}+Score_{\mathrm{SC}}+Score_{\mathrm{FC}}.$$

Regarding the total score, a patient or his doctor can make some judgments that if he reacts more slowly than before, and if he needs to see the doctor for further professional diagnosis, such as WAIS-R, WAMS-R and HDS.

EXPERIMENTS AND EVALUATION

We invited 28 people as the experiment objects. In order to observe the distribution of reacting capability in age space, three age groups of people were collected: the young who were under 20, the middle aged who were between 20 and 40, and the elderly who were over 40. Also, to find out what is the variance of the reacting capability as people get tired, we acquired the data twice, at 9:00 o'clock when people are full of energy and at 13:00 just when people get tired after 4 hours' work.

To eliminate the negative influence of unfamiliarity with the system, before we get started to acquire the data, the objects were asked to play the game with computer for no less than 2 h as necessary training. In actual experiment, these three groups of objects played the game freely, but four cases of SL, SC, FL, FC are required to be covered. Their scores were recorded automatically. The achievement of each group is listed in Table 2. According to the scoring rule noticed in last section, one correct gesture in SL case can win one score, in SC and FL case, can win two, and in FC case, can win three. The final scores are the average score of each group that had been normalized to the range of (0,100).

Table 2 Experiment result

Item		Score		
		The young (5)	The middle-aged (20)	The elderly (3)
SL	9:00	48	182	47
	13:00	49	182	46
SC	9:00	90	336	44
	13:00	90	330	40
FL	9:00	92	342	46
	13:00	90	338	48
FC	9:00	117	477	54
	13:00	111	450	33
Final	9:00	86.75	83.56	79.59
	13:00	85.00	81.25	69.59
Variance percentage (%)		2.02	2.76	12.56

Note: The numbers of 5, 20 and 3 in the brackets represent the numbers of people in those groups

From Table 2, we can make some conclusions: (1) People's reacting capability decreased with age. (2) People's reacting capability decreased when they are tired. Meanwhile, we deem that the sum of four phases' scores represents each group's level, and that the difference of each group's level between 9:00 and 13:00 is shown in Fig.8. From the chart, we can see that it is the elderly who are affected most by fatigue.

As we know, knowledge we get from the Table 2 and Fig.8 are identical to the fact we can observe in life. So our game system can be utilized as a tool for reacting capability evaluation.

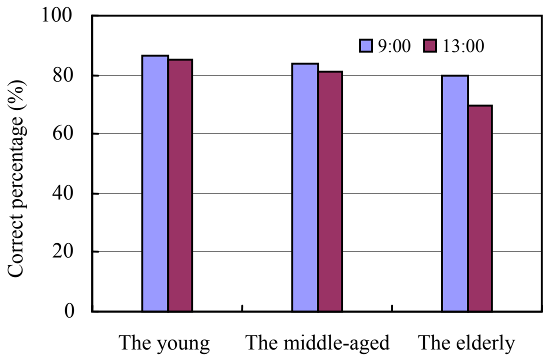

Fig.8 The variance of reacting capability

CONCLUSION AND FUTURE WORK

This paper presents a dementia examination system helping people to make an initial judgment about his mental status before seeing the doctor and undergoing expensive medical check. The system effectively reflects people's reacting capability. Importantly, it is easy to operate, people can obtain their mental status by being a game player. It is a preferable choice for the elderly and mental worker.

However, there are still some significant tasks that need to be continued. The first is how to arrange the experiment step adequately in order to mine completely the relationship between people's reacting capability and potential dementia. The second is to improve the robustness of the moving gesture recognition problem. We expect to realize those targets in future work.

References

Almkvist, O., Basun, H., Backman, L., Herlitz, A., Lannfelt, L., Small, B., Viitanen, M., Wahlund, L.O., Winblad, B., 1998. Mild cognitive impairment: an early stage of Alzheimer's disease. *J. Neural. Transm. Suppl.*, **54**:21-29.

Ara, V., Monson, H., 1999. Face Recognition Using an Embedded HMM. IEEE Internatona1 Conference on Audio Video Biometric Based Person Authentication. Washington D.C.

Chen, Y.Q., Gao, W., Wang, Z.Q., Yang, C.S., Jiang, D.L., 2002. Text to avatar in multimodal human computer interface. *Asia-Pacific Human Computer Interface (APCHI2002)*, **2**:636-643.

Chen, Y.Q., Gao, W., Fang, G.L., Wang, Z.Q., 2003. CSLDS: Chinese Sign Language Dialog System. IEEE International Workshop on Analysis and Modelling of Faces and Gestures (held in conjunction with ICCV-2003). Nice, France, p.236-237.

Francis, P.T., Palmer, A.M., Snape, M., Wilcock, G.K., 1999. The cholinergic hypothesis of Alzheimer's disease: A review of progress. *J. Neurol. Neurosurg. Psychiatry*, **66**:137-147.

Ge, C.B., Chen, Y.Q., Yang, C.S., Yin, B.C., Gao, W., 2005. Motion Retargeting for Hand Sign Language. Proceeding of WSCG.

Geldmacher, D.S., Whitehouse, P.J., 1996. Evaluation of dementia. *New England J. Medicine*, **335**(5):330-336. [doi:10.1056/NEJM199608013350507]

Guo, Q.H., Qin, Z., Lü, C.Z., 2000. A review of cognitive capability scale for Alzheimer's disease. *Chinese Journal of Neurology*, **33**(3):179-182 (in Chinese).

Kawas, C.H., 2003. Clinical practice. Early Alzheimer's disease. *New England J. Medicine*, **349**(11):1056-1063. [doi:10.1056/NEJMcp022295]

Knopman, D.S., DeKosky, S.T., Cummings, J.L., Chui, H., Corey-Bloom, J., Relkin, N., Small, G.W., Miller, B., Stevens, J.C., 2001. Practice parameter: diagnosis of dementia (an evidence-based review). *Neurology*, **56**:1143-1153.

Kuo, S.S., Agazzi, O., 1994. Keyword spotting in poorly printed documents using pseudo 2D Hidden Markov Models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **16**(8):842-848. [doi:10.1109/34.308482]

Rabiner, L., 1989. A Tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of IEEE*, **77**(2):257-286. [doi:10.1109/5.18626]

Yang, C.S., Wang, Z.Q., Gao, W., Chen, Y.Q., 2003. How to build the skeleton of the individual virtual human model. *Journal of Computer Aided Design & Computer Graphics*, **16**(1):67-73 (in Chinese).

Zhang, L.G., Chen, Y.Q., Fang, G.L., Chen, X.L., Gao, W., 2004. A Vision-Based Sign Language Recognition System Using Tied-Mixture Density HMM. ACM Sixth International Conference on Multimodal Interfaces (ICMI'2004). State College, Pennsylvania, USA, p.198-204.