Avatar-based Intelligent Navigation for Online Shanghai Expo

Mingmin Zhang[1], Nan Xiang[1], Kangkan Wang[1], Zhigeng Pan[1], Haibo Yuan [2]and Houquan Liu [2]

[1]State Key Lab of CAD&CG, Zhejiang University, Hangzhou, 310027, China
[2]China University of Mining and Technology, Xuzhou, 221008, China

Abstract—**Shanghai Expo 2010 is an international event which will demonstrate the scientific, economic and cultural advance of the exhibiting countries. The local government launches a project named on-line Shanghai Expo in order to make the event accessible to people through Internet. Virtual reality technologies, especial intelligent virtual environment (IVE), are exploited to support the Expo. IVE, based on semantic, is the integration of Virtual Environment and Artificial Intelligence. It can improve the interaction between avatar and the virtual world by giving the objects in the environment semantic information. In this paper we employ intelligent virtual environment to build a virtual environment of realistic behavior and intelligent navigation.**

Index Terms—**Virtual Reality, Intelligent Virtual Environment, semantic information, avatar based interaction.**

I.   INTRODUCTION

The World Expo has attained increasing prominence as grand events for economic, scientific, technological and cultural exchanges, serving as an important platform for displaying historical experience, exchanging innovative ideas, demonstrating esprit de corps and looking to the future. Expo 2010 Shanghai China will also be a grand international gathering. On the one hand, it shall endeavor to attract about 200 nations and international organizations, as well as 70 million visitors from home and abroad, to take part in the exhibition, ensuring the widest possible participation in the history of the World Expositions. On the other hand, it will put Expo 2010 Shanghai China in a global perspective and do the best to encourage the participation and gain the understanding and support of various countries and peoples, in order to turn Expo 2010 Shanghai China into a happy reunion of people from all over the world[1]. So, it is important to construct an online Expo to help the people of the world on visiting this show.

Intelligent Virtual Environments (IVE), formed by the increasing overlap of the technologies involved in 3D real-time interactive graphics environments and Artificial Intelligence/Artificial Life technologies. We utilize semantic information to construct IVEs.

Semantics in a virtual scene can help us record and analyze the avatar behavior, which is used to implement intelligent navigation. More guides can be supplied to users while navigating in virtual environments. Igor E. Tom describes high-level control of virtual human's behavior and indicates that behavior primitive library should be created in the form of appropriate database. In addition, we can acquire more real and smooth interaction with virtual environments using avatars. The other purpose for studying avatars is to support decision-making system. Lydie E. deals with the design of training system to support decision-making in the preparation and the management of maintenance interventions in high-risk industries namely SEVESO sites[2].

In order to make the interaction effectively, we adopt avatars to connect the users and virtual environment. The adoption of virtual characters generally has two main functions: first, the virtual characters can draw huge attractions from the audience and bring a public recognizing to the museum; second, the virtual characters can make the visitor gain authentic experience about the situations which might not be existed any more [3]. Virtual characters can play important role in virtual systems because it can enhance communication efficiency via manners of interpersonal communication by well representing the information as well as improving the general publicity [4, 5]. Pan examines the categories of virtual characters by studying the story genres and identifying the potential resources for creating new virtual characters: the traditional and contemporary cultural resources [6].

The avatars can be customized in order to enhance the realistic behavior of avatar (Virtual Human). Recently, personalization has become a topic of interest because of its ability to provide a better user experience in this era of virtual reality. Studies report that personalization positively impacts human-technology relationships [7]. For example, online services that provide information and knowledge tailored to individuals have more satisfied customers [8]. However, existing studies about personalization have largely focused on the parameters setting for users' custom. Our project integrates personality with avatar construction. Users can manipulate an avatar with their own face. We employ a photo to reconstruct the user's 3D face model to implement the customization. There are two steps: first, extracting the characteristics of the user from his photo; second, reconstructing face model according to extracted characteristics.

ASM algorithm (Active Shape Model) is based on statistics [9-14]. For a new input image, it first detects the face region, and then in a local area searches for the feature points through the rotation and scaling operation to find out the actual feature points which are closest to the provided points in the template. For the grid deformation part, we employ the RBF (Radial Basis Function) -based interpolation method.

In the following part of this paper, detailed descriptions

about architecture and key technologies of Shanghai Online Expo including IVE construction and personalization are given in Section 2. Section 3 is on the avatar's behaviors recording and analyzing for intelligent navigation. The structure of recording and the method for behaviors analyzing will be presented. Experimental results that have been done in the project of Online Shanghai Expo will be shown in the end, followed by the conclusion and acknowledgement.

II.   FRAMEWORK OF ONLINE EXPO

Expo 2010 Shanghai China will centre on innovation and interaction. Innovation is the soul, while cultural interaction is

(a)   Theme pavilions

(b)   Harmony square

Fig. 1. View of Shanghai Expo

an important mission of the World Expositions. In the new era, Expo 2010 Shanghai China will contribute to human-centered development, scientific and technological innovation, cultural diversity and win-win cooperation for a better future, thus composing a melody with the key notes of highlighting innovation and interaction in the new century[1].

As an aspect of innovation, providing a better user experience is a very important design principle we followed during the project.   A better user experience evolves personalization, realistic environment, interactive interface and rich contents.

2.1 Architecture of the on-line Expo

The architecture of the online expo is shown in Fig. 2. It involves four parts: the virtual environment, avatar based interface, user information database and navigation component.
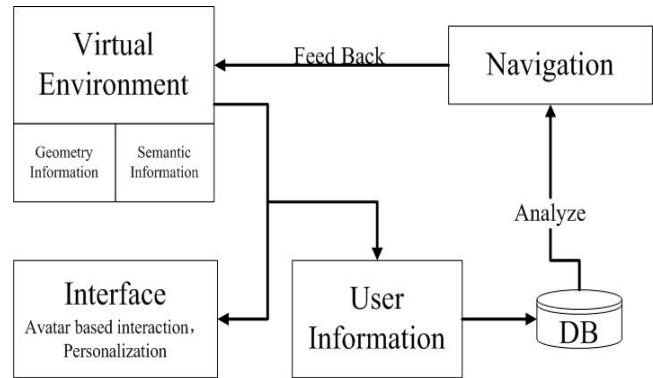

Fig. 2. Process of user's information extracting and analyzing

2.2 Key techniques

(1) Semantic information
Semantic refers to the potential meanings of an entity, which is different from its appearance features that can be seen by people. The semantic information also includes the relations between the entity and its surrounding environment. Semantics in a virtual scene refer to information implied in virtual entities which are formed by computers, as well as related information with others. For example, a virtual glass-cup contains the following semantic information: once it is over capacity then water will overflow, if it falls on the ground it will be broken, and so on. If the cup is placed on a table and when the table is moved away, it will drop from the table. This is the semantic information between the cup and the table.

Assuming that user a need to take some operation on the object b in virtual scene, the rule is that the distance between a and b is equal to or less than a distance of c in six directions. If the condition is met object b would be active and accept operation. Otherwise any operation will not cause reaction in the scene. The rule is described as the following logic expression

$$\forall a \forall b \exists c (user(a) \wedge object(b) \wedge (dist(a,b) < c)) \Rightarrow active(b)$$

In this virtual scene, there is a user a and an object b which can be active to accept certain interaction task. The statement is: only meaningful task on objects will be treated by the system. For example, the wall of a virtual building needn't any interaction with avatar. So the wall's information is just geometric data. But if there is a door in the wall, the semantics information will be necessary. When an avatar is far away, the door is closed. As soon as the avatar walking to the door and being close enough to the door, it will open automatically. In a

word, not all the objects can be interacted with. The term *dist (a, b)* is to calculate the distance between *a* and *b*, while *active (b)* means b is active.

(2) Personalization

ASM is a statistic algorithm and the existed training templates, usually Caucasian, are quite different from Asian. For this reason we use the general face template and train it in our algorithm. There are 77 feature points in our algorithm, with 5 points corresponding to the eyebrows, 5 to the eyes, 12 to the nose, 11 to the mouth and 34 to the outline of the face. The coordinates of these 77 points are listed as a vector.

Through the shape alignment, the feature points in the standard face have been located in a specific facial photograph. Then next step is the local fine-tuning of the gray-scale image information in the specific face. Searching for image borders can make the feature points align with the gray-scale border of the image.

Different person has different hair style. This becomes a problem to ASM for the training template varies depending on different hairstyle. In the forehead, the ASM algorithm can't achieve satisfactory results. Therefore current ASM algorithms do not include feature points in the forehead. We found that most of the human faces had a proportion so that each face could be divided into the upper, middle and lower part in a certain proportion. The additional points can be linear interpolated directly in accordance with the proportion relation and combined with the restriction in the face (as shown in Fig. 3).
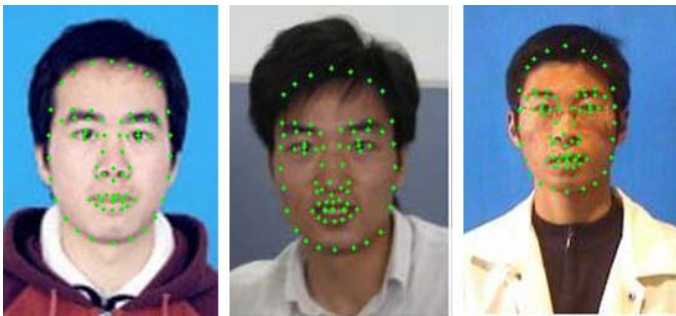

Fig. 3. Feature points detection

In order to reduce the burden of adjustment, the initial face model should be a standard one, which must be a good neutral model for men and women, middle age, with a neutral facial expression. We use FaceGen to generate an initial Asian model, which is shown in Fig. 4 and Fig. 5. Our method generates a specific 3D human face model through the transformation from the neutral model to a specific one, and then creates a specific 3D face by realistic texture mapping to the specific 3D human face model. In the amending process from the neutral face model to the particular one, it is necessary to carry out two transformation operations. Firstly we use overall transformation to the neutral face model to complete the amending of the overall face outlines. The face model should fit the location of the specific face shape and the main organs. Then we do the local transformation to the whole neutral face model in order to amend the shape and the size of the brows, eyes, mouth, nose based on a specific person, and mark the specific features in the neutral face model. The current method employs the standard model as a general human face model,

matching the points in the frontal human face photos to adjust a number of key points to obtain a specific face model.

The feature matching makes the feature points in the 3D model as close as possible to those in the 2D image by deforming the feature points in the 3D model to the 2D feature points. After texture mapping, the 3D standard model turns to a specific 3D model of a person.
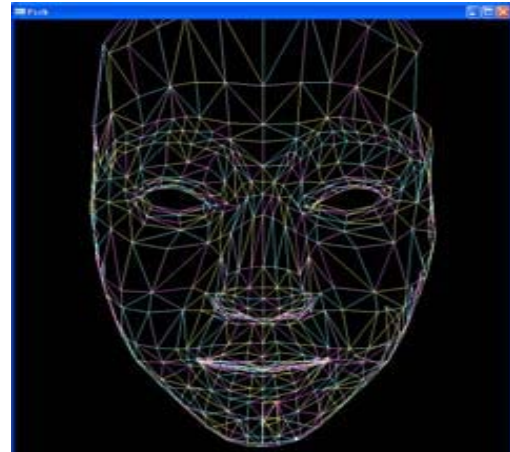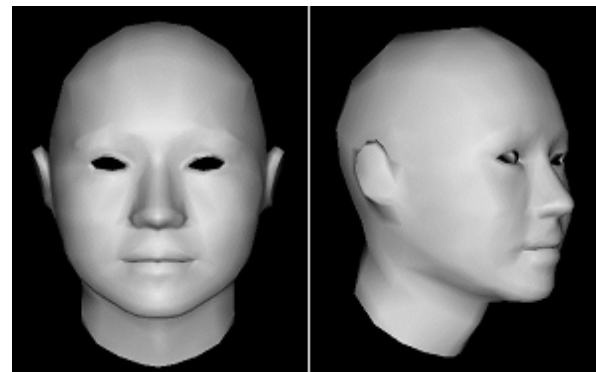

Fig. 4. Neutral face model

Fig. 5. Neutral human face entity model

Fig. 6. Personalization result

III.   IMPLEMENTATION OF INTELLIGENT NAVIGATION

3.1 Avatar-based interface

We employ an avatar-based interaction, which is different from the traditional interactive methods, to enhance the immersive felling of users to the on-line Expo environment. An avatar is an embodiment of a user in the virtual world which has fundamental differences with any cartoon figure in printing media, because it can bring the audience to the real situation. It implements the task of event trigger during the interaction progress. Integrated with semantic objects in virtual environments, the avatar can "walk" to the objects which users are interested in, and "observe" them to trigger an interaction process. Furthermore it can receive the feedback generated by the task implement part. So it is an important mediation for the audience to understand the content which the virtual environment would like to exhibit. In Online Shanghai Expo visitors could walk around the displayed areas sections by sections to learn the information under the aids of catalogues, text panels, extended labels, etc. If they participated in a group, they can communicate with each other by typing words.

User-defined virtual characters are also provided in this system. The user can pick up or define a new virtual character to explore the experience of dialoguing with other virtual characters. Avatar-based virtual environment can increase social presence and interpersonal trust among communicators in the process of collaborations leading by virtual characters.

3.2 Intelligent Navigation

To make the interaction smarter, avatar behavior recording and analyzing are necessary. In the Section 2 the main model for 3D operation, semantics of virtual objects have been introduced. This part will give the design for behavior recording and analyzing: what the user has done in virtual scene and what we can get from his/her operation record.
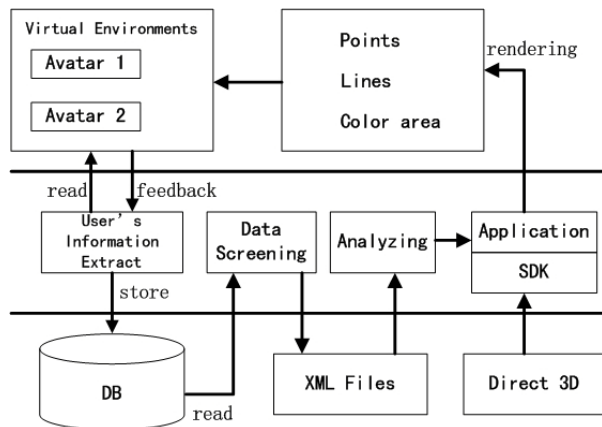

Fig. 7. Process of user's information extracting and analyzing

Fig. 7 shows the main process of user's information extracting and analyzing. Firstly, the extracting module gets information package from virtual environment within a certain timeslice. All kinds of data are stored in the DB. And before analyzing, the data screening is necessary. Different analyzing

purposes determine what kinds of data are to be read or used. The results are stored in XML files and it is easy to read. The analyzing results will be sent to application program which is developed on 3D Game SDK. There mainly are three kinds of methods for representing the analyzing results: points, lines and color area. Points and lines are used to describe the avatar's path and current location. The color area is used for showing the user flow of each scene.

Information package is necessary for the communication between system and virtual scene. The design of package is as following:

Fig.8. Package structure

Once the extracting module receives a command to get a package, the above information will be packaged and stored into DB.

3.3 Recording module

In the virtual on-line Expo the whole scene is divided into four main parts: inside, outside, first floor and second floor. As the number of textures and maps in the scene is very large, each part has its own folder to store textures and maps. Once an avatar enters a scene from another scene, the system needs to reload the scene files.

For example, <order1>13241<order1>, this record means during the avatar's first login he/she has visited scene 1, scene 3, scene 2, scene 4 and back to scene 1 in order.

In the "Functions" model, there are six functions to implement different tasks. "*Monitor ()*" is to detect whether
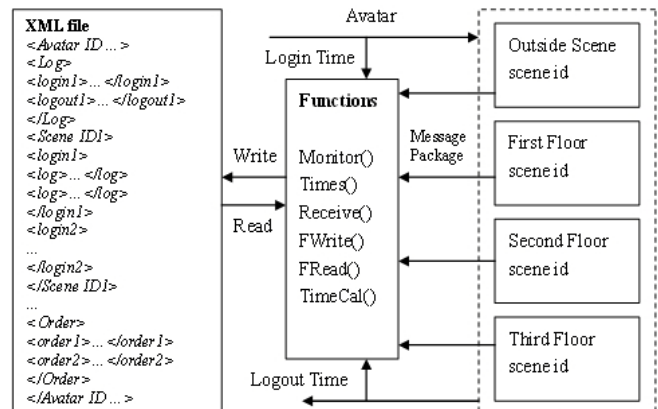

Fig. 9. Model of user's behaviors recording

there is an avatar entering. Once an avatar enters the system, the function "*Times ()*" will be used to find the entering time. After the avatar enters a sub-scene, message packages sent by the sub-scene will be received by "*Receive ()*". A message package contains the following information: sub-scene id, login time and some variables that represent different kinds of operations. The received information is written into XML file by "*FWrite ()*" and would be read by "*FRead ()*" when necessary. How long the avatar stays in a sub-scene will be calculated by "*TimeCal ()*".

As the record <order1>13241<order1>, the time that avatar stays in sub-scene 3 is the result of login time of sub-scene 2 minus that of sub-scene 3. That means it just needs to get the result from the latter minus the former in a neighboring sub-scenes. And we will know how long the avatar stays in the former sub-scene. But if the avatar enters a sub-scene more than once, we should add them again and again. And the time for the last sub-scene is equal to the logout time minus its entering time.

3.4 Experimental analyzing

After extracting information from virtual scene, the analyzing is the next important work. For example, if you want to know the path you have walked through in the previous half hour, the first step is to read related information. In the case, the only information that is required is the coordinate of the avatar. If the system gets information package every five minutes, there should be seven records. Point is used to describe avatar's location in each record. We suppose that the path between two adjacent points is a straight line. So the path will be showed in Fig. 10.
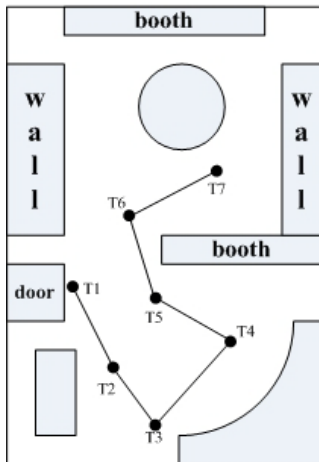

Fig. 10. Avatar's path in past half hour

To get the user flow of each scene color area is used to illustrate the different situation. For example there are four kinds of colors to represent different user flow. If there are more than eighty users visiting the scene during the past half hour, the deep blue is used. And the blue is for sixty to eighty, light blue is for thirty to sixty, finally the gray is for less than thirty. Analyzing begins with searching the records in DB according to the scene ID and time, because there will be a corresponding record as long as an user logs in. The experimental situation is showed in Fig. 11.

According to the information records from the users, we build an intelligent interactive interface to make the system smart, just using users' behaviors simply. The system can recommend the most popular scene to the users.
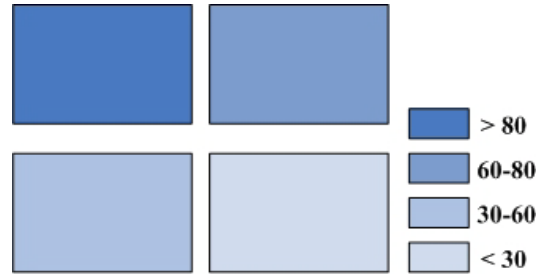

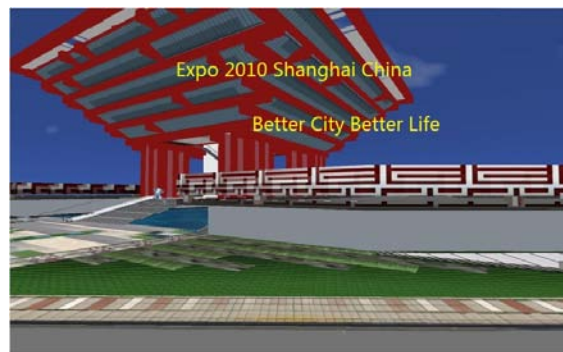Fig. 11. Representing user flow for each scene with different colors


Fig. 12. Login interface.(Color Plate 6)

Fig. 13. Personalized face model (Color Plate 7)

IV.   EXPERIMENT RESULTS

Based on the above study, a distributed Online Shanghai Expo prototype system has been developed by the researchers from three universities and five companies as shown in Fig. 12. User can log in the system and navigate the virtual scene by controlling an avatar which can be created based on the photos from the user. Employing ASM and RBF, we construct a personalized face model as shown in Fig. 13. The bottom image is the original photo, and the top is the customized face model, created from the bottom image.

User can also "talk" with other avatars that are in the same virtual world as shown in Fig. 14. Furthermore, in the future work they can say hello to their friends using microphone connected to the computer. Another instance is for intelligent perception as shown in Fig. 14(a) and 14(b): if the avatar is not close enough to the virtual TV, the light and TV will not be turned on, if the distance between avatar and TV is less than the preset length the light will turn on and the video will be played in TV. In this experiment the semantic rule is used to control the interaction among the avatar, light and TV. The interaction will be triggered only when the distance between these objects is less than a defined value. Fig.14(c) shows the intelligent interaction interface for recommending scene to users. Users could choose to accept the help of the system or not. According
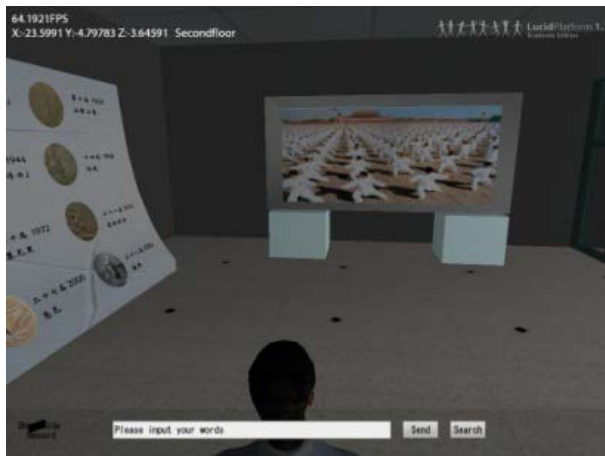
(a)   the avatar is not close enough

(b)   the light and TV are turned on

(c)   Virtual scene with interactive window

(d)   Choose "no" to cancel navigation

Fig. 14. Intelligent navigation (Color Plate 8)

to the user experience records in the database, the system provides a better route for the user.

V.   CONCLUSIONS

Semantic-based interactive 3D technology has supplied an effective solution for intelligence in virtual environments. In our work the semantic information of virtual objects and interactive operations have been employed to record and analyze the avatar's behavior. Furthermore, avatars are employed to implement intelligent interaction with users. It can enhance the immerse feeling of users in Online Shanghai Expo. However there are still several problems that need to be improved. The library of semantics is not big enough for most interaction. And the analysis results are not enough for the avatar's operations in virtual environments.

As a result, future work will focus on enriching the semantics library and analyzing the complex actions. We hope to obtain the mechanism of semantic information automatically. Therefore further study on the analyzing of complex operations will be more useful and desirable.

VI.   ACKNOWLEDGMENT

REFERENCES

[1] "http://en.expo2010.cn/a/20081116/000004.htm."

[2] L. Edward, D. Lourdeaux, D. Lenne, J. Barthes, and J. Burkhardt. "Modelling autonomous virtual agent behaviours in a virtual environment for risk," *International Journal of Virtual Reality*, 7(3), pp. 13–22, 2008.

[3] P. Shao, W. Liao, and Z. Pan. "Adopting Virtual Characters in virtual systems from the perspective of communication studies," *Transactions on Edutainment II*, pp. 70-89, 2009.

[4] S. Gbel, I. Iurgel, M. Rsler, F. Hülsken, and C. Eckes. "Design and narrative structure for the virtual human scenarios," *International Journal of Virtual Reality*, 6(4), pp. 1-10, 2007.

[5] F. Hülsken, C. Eckes, R. Kuck, J. Unterberg, and S. Jrg. "Modeling and animating virtual humans for real-time applications," *International Journal of Virtual Reality*, 6(4), pp. 11-20, 2007.

[6] Z. Pan, W. Chen, M. Zhang, J. Liu, and G. Wu. "Virtual reality in the digital olympic museum," *IEEE Computer Graphics and Applications*, 29(5), pp. 91-95, 2009.

[7] J. Sung, R. Grinter, and H. Christensen. "Pimp my roomba: designing for personalization," Apr.2009 *proceedings of the 27th international conference on Human factors in computing systems. ACM*, pp. 193-196.

[8] G. Adomavicius, and A. Tuzhilin. "Personalization technologies: a process-oriented perspective," *Communications of the ACM*, 48(10), pp. 90, 2005.

[9] A. Hyv rinen, and P. Hoyer. "Emergence of phase-and shift-invariant features by decomposition of natural images into independent feature subspaces," *Neural Computation*, 12(7), pp. 1705-1720, 2000.

[10] T. Cootes. "An introduction to active shape models," *Image Processing and Analysis*, pp. 223–248, 2000.

[11] S. Milborrow, and F. Nicolls. "Locating facial features with an extended active shape model," *Computer Vision–ECCV 2008*, pp. 504-513, 2008.

[12] E. Lee, J. Kwon, and I. Lee. "Caricature video," *Computer Animation and Virtual Worlds*, 18(4-5), pp. 279-288, 2007.

[13] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. Salesin. "Synthesizing realistic facial expressions from photographs," *Jul .2005,ACM SIGGRAPH 2005 Cources*, pp. 19.

[14] E. Akleman. "Making caricatures with morphing," *Jan .1997 ACM SIGGRAPH 97 Visual Proceedings : Theart and Interdiscip Linary Programs of SIGGRAPH'97* , pp. 145.

Mingmin Zhang is an associate professor of Computer and Engineering Department, Zhejiang University. She got the Bachelor degree from Computer Science Dept, Nanjing University in 1990, and the Master and PhD degree from Computer Science and Engineering Department, Zhejiang University in 1995 and 2008.

She has published more than 40 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.

Nan Xiang is currently a doctoral candidate in the State Key Lab of CAD&CG, Zhejiang University. He received his Bachelor's degree and Master's degree in Computer Science and Technology from Xi'an Jiaotong University in 2006 and Chongqing Post and Telecommunication University in 2009, respectively. His research interests include character animation, computer game and affective computing.
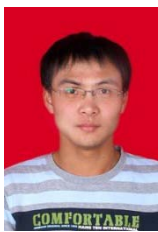
Wang kangkan received the Bachelor degree in Computer Science from Northwestern Polytechnical University.,In June 2009.

Since February 2009, he has been participating in the research project of Shanghai World Expo. In the project, he has been dealing with related research on the semantics-based interaction in virtual environment. In September 2009, he began the PhD in CAD&CG laboratory of Zhejiang University.

Zhigeng Pan was born in 1965 in Jaingsu Province, and he received his Bachelor Degree and Master Degree from the Computer Science Department in 1987 and 1990 from Nanjing University respectively and Ph.D Degree in 1993 from Zhejiang University. Since 1993, he has been working at the State Key Lab of CAD&CG, Zhejiang University.

He has published more than 100 papers on international journals, national journals and international conferences. His research interests include distributed graphics, virtual reality, multimedia, digital entertainment.

Prof. Pan 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, Co-EiC of LNCS Transactions on Edutainment. 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, et al. He is the Director of the Virtual Reality Committee, China Society of Image and Graphics; He is on the Steering Committee member of VRCAI series conference. He is the organizing committee chair of VRAI'2002, ICIG'2004. He is the program co-chair of EGMM'2004(Eurographics workshop on Multimedia), the program co-chair of Edutainment'2006, conference co-chair of ICAT'2006 and Edutainment'2008.