Skills Assessment of Users in Medical Training Based on Virtual Reality Using Bayesian Networks

Ronei M. Moraes[1], Liliane S. Machado[1], and Leandro C. Souza[2]

[1] Federal University of Paraíba, João Pessoa/PB, Brazil
ronei@de.ufpb.br
[2] Federal University of Pernambuco, Recife/PE, Brazil
liliane@di.ufpb.br

Abstract. Virtual reality allows the development of digital environments that can explore users' senses to provide realistic and immersive experiences. When used for training purposes, interaction data can be used to verify users skills. In order to do that, intelligent methodologies must be coupled to the simulations to classify users´ skills into N a priori defined classes of expertise. To reach that, models based on intelligent methodologies are composed from data provided by experts. However, online Single User's Assessment System (SUAS) for training must have low complexity algorithms to do not compromise the performance of the simulator. Several approaches to perform it have been proposed. In this paper, it is made an analysis of performance of SUAS based on a Bayesian Network and also a comparison between that SUAS and another methodology based on Classical Bayes Rule.

Keywords: Medical Training, User's Assessment, Bayesian Networks, Virtual Reality.

1 Introduction

Virtual Reality (VR) environments can provide realistic systems for several areas and have been used since a long time [1]. In such immersive and interactive environments, users perform tasks that simulate real situations. Interaction data feed the system and are processed in real time to generate feedback for users, as new points of view, force feedback and sound effects. This data can also be collected to monitor users´ actions in order to analyze how tasks are accomplished and classify users performance. This is particularly interesting when the virtual environments are used for training purposes.

Researches on training assessment for simulators based on VR has less than 20 years old and the methodologies for assessment make possible to know users' performance during the training to analyze if they are prepared to perform the procedure in real situations. Probably the pioneer works for Single User's Assessment System (SUAS) were proposed by Dinsmore et al. [3] using a quiz to assess users in the identification of subcutaneous tumors in a training system based on VR. The quiz was composed by questions related to the diagnosis and hardness of tumor. Other research group [16] created a minimally invasive system in which each task could be programmed for different difficulty levels. Performance data of each user could be saved to post analysis (offline) by an expert or using statistical methods [15].

L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 805–812, 2012.

Since 90's, several assessment methods were proposed [8,13], mainly for medical training. With continuous advances on computers performance, SUAS evolved too. Nowadays, a SUAS must continuously monitor all users interactions on VR environment and compare their performance with pre-defined expert's classes of performance to recognize users level of training. Basically, there are two types of SUAS: off-line and on-line.

Off-line SUAS can be defined as methods which can be or not coupled to VR systems, whose assessment results are provided some time (which can be minutes, hours or days) after the end of the VR-based training [15]. On the other hand, on-line SUAS must coupled to the training system based on VR [10] and should be able to collect interaction data to provide a result of performance in lower than one second after the end of the simulation [12]. It occurs by comparing a model composed by users' interaction data with classes of performance previously defined from experts knowledge. A SUAS must be able to monitor user's interactions with the VR simulator by variables such as spatial position (of user and of interaction tools used in the VR simulation), touch force and resistance, user's angle of visualization, sound, smell, sense of temperature, velocity and acceleration of interactions and felt shapes. All the information are sent to the SUAS which analyzes the data and emits, at the end of the training, an assessment report about user's performance according pre-defined classes of performance (Figure 1). An on-line SUAS must have low complexity to does not compromise VR simulations performance, but it must have high accuracy to does not compromise the assessment and they are normally based on pattern recognition techniques. Then, several methodologies developed for SUAS can be potentially applied to other research areas.

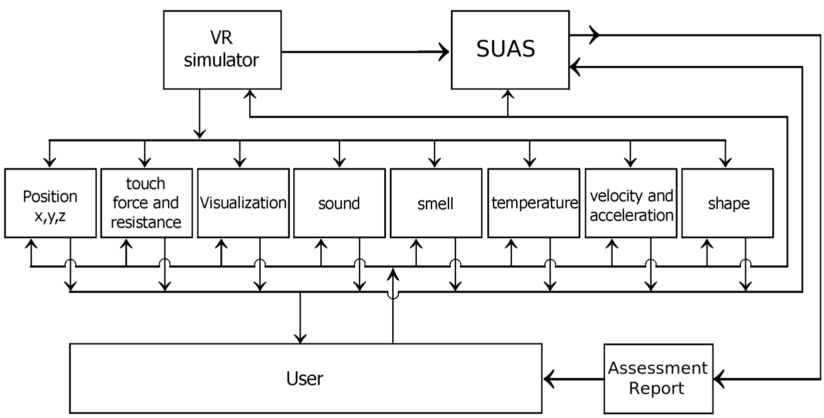

Fig. 1. Diagram of a training system based on VR with a SUAS coupled on it [9]

An approach based on a Bayesian Network for an online SUAS was proposed by [11]. In the present paper we provide an analysis of performance of that approach and also a comparison between that approach and another one based on Classical Bayes Rule [9,12].

In medical training assessment, is necessary to find the better SUAS to assess users with high accuracy. Since patients' life or their health conditions can depend on physicians skills, a good assessment method is important to provide feedback about training status. However, the choice of a methodology will depend on the features of the training simulator that can provide qualitative and/or quantitative data as input for the assessment. Since the kind of input data is known, models can be composed and integrated into a SUAS. For performance reasons, supervised methods are the most used in SUAS, but an exhaustive comparison them was not found in the literature. Then, the main contribution of this paper is the presentation of a methodology for SUAS bases on Bayesian Networks and its comparison with another methodology based on Classical Bayes Rule. A simulator for training of a medical procedure is presented as an example of application of the SUAS.

2 Theoretical Aspects

This section presents theoretical aspects of the SUAS based on Bayesian Network. For reader's better understanding, it is presented a short review about the Classical Bayes Rule and in the following, is presented the Bayesian Network.

2.1 Classical Bayes Rule

Formally, let be the classes of performance in space of decision $\Omega=\{1,...,M\}$ where M is the total number of classes of performance. Let be w_i, $i \in \Omega$ the class of performance for an user. The Classical Bayes Rule (CBR) [9,12] computes conditional class probabilities and then predict the most probable class of a vector of training data $\boldsymbol{X}$, according to sample data D, where $\boldsymbol{X}$ is a vector with n features obtained when a training is performed, i.e. $\boldsymbol{X}=\{X_1, X_2, \ldots, X_n\}$. Using the Bayes Theorem:

$$P(w_i \mid \boldsymbol{X}) = [P(\boldsymbol{X} \mid w_i)\, P(w_i)] \,/\, P(\boldsymbol{X}) \Leftrightarrow P(w_i \mid X_1, X_2, \ldots, X_n) = \\ = [P(X_1, X_2, \ldots, X_n \setminus w_i)\, P(w_i)] \,/\, P(\boldsymbol{X}) \quad (1)$$

However, as $P(\boldsymbol{X})$ is the same for all classes w_i, then it is not relevant for data classification. Then, the equation (1) can be expressed by:

$$P(w_i \mid X_1, X_2, \ldots, X_n) = P(w_i)\, P(X_1, X_2, \ldots, X_n \mid w_i) \quad (2)$$

Then, the assessment rule for CBR is done by:

$$X \in w_i \;\; if \;\; P(w_i \mid X_1, X_2, \ldots, X_n) > P(w_j \mid X_1, X_2, \ldots, X_n) \\ \text{for all } i \neq j \text{ and } i, j \in \Omega \quad (3)$$

2.2 Bayesian Network

Formally, a Bayesian network is defined as directed acyclic graphs, denoted by G and a probabilistic distribution denoted by P. The graph G is a set of nodes and oriented arcs, where nodes represent variables in process and oriented arcs encode conditional dependencies between that variables [14]. The dependencies are modeled by specific conditional probabilistic distributions [6].

Several kinds of Bayesian networks can be found in literature [2]. They can differ on their graph structure and types of dependencies relationship which they can model. According to statistical relationship between variables which describe that process a specific Bayesian network should be chosen. This is critical, because it changes the final results. The General Bayesian Network (GBN) is a generalized form of Bayesian networks, which allows nodes to form an arbitrary graph [2]. Another important characteristic is that each child node cannot be connected to the final classes of assessment and the dependencies between nodes can adjust itself to real dependencies. Thus, it is possible to verify dependencies between variables during network modeling and put them in structure nodes of GBN, which did not occur in other structures [5].

Formally, let be the same M classes of performance, w_j, $i \in \Omega$ the class of performance for a user and X_k, $1 \leq k \leq n$, represents a node in GBN with n as the number of nodes in a graph. The joint probability distribution in GBN for an event is done by:

$$P(X_1, X_2, \ldots, X_n) = \Pi^n{}_{k=1} \quad P(X_n \mid X_{n-1}, X_{n-2}, \ldots, X_1) \tag{4}$$

where $P(X_1, X_2, \ldots, X_n)$ is the joint probability distribution and $P(X_n \mid X_{n-1}, X_{n-2}, \ldots, X_1)$ is the conditional probability of X_n conditioned by its predecessor nodes $X_{n-1}, X_{n-2}, \ldots, X_1$.

The probability nodes are associated to probability distribution, which can be different for each node. For example, a node A can have a Gaussian distribution and a node B, which depends on A, can have a bivariate Gaussian distribution. The structure of GBN is learned from data, finding dependencies among nodes, as well as the parameters of conditional probabilities. Scores are used to help estimate the final structure of GBN for each class of assessment. In a first moment a network is created with all independent nodes and an initial score is calculated. Next, all combinations are searched and an arc is designed between two nodes for which an increment of initial score is obtained. Then, the parameters for that nodes set are re-estimated using linear regression. This cycle is repeated until total network score is less than a predetermined value or a fixed number of cycles [11].

3 The Bone Marrow Harvest Simulator

A bone marrow harvest simulator was developed to allow training of the bone marrow harvest for transplant [7]. The procedure is composed by three steps which were considered in the development of the simulator. The first step is related to the observation of the anatomy, the second to the palpation of the patient pelvic region, and the third to the puncture process to harvest bone marrow. Since the main activity of the procedure is related to the third step, the simulator was designed in details to represent this step. Also, a haptic device was used to provide realism and co-relation between real and virtual environments.

Since this simulator included a rigorous process of design, a framework called CyberMed [8] was used to decrease the development effort and speed up this process. The framework is expansible and several methodologies for a SUAS can be included in its functions. This makes possible to observe the performance and the efficiency of the assessment methodologies in a simulated training. In this case, the haptic interaction data of the puncture step is acquired and used to assess the user in order to indicate their dexterity a ability to perform the procedure.

In a usual situation, a physician should calibrate the SUAS previously, according *M* classes of performance defined, where *M* may be any integer number greater than one. The calibration process consists on execute several times the procedure in the simulator and to label each execution according to classes of performance. This calibration occurs in a simulation previously designed to capture physician interactions. This simulation is similar to the final simulator but the physician interaction data is collected and labeled for each execution of the procedure. The number of classes of performance is defined by the physician, e.g. as $M=3$: 1) correct procedures, 2) acceptable procedures, 3) badly executed procedures. So, the classes of performance for a trainee could mean: "you are well qualified", "you need some training yet", "you need more training".

A SUAS based on GBN was implemented in Cybermed, expanding the framework capabilities. This implementation added a new class in the CybAssess class. In this implementation, the default value for network score is 10^{-4} and the fixed number of cycles is 100.

4 Results

To test the GBN method, four classes of performance were simulated from seven different Gaussian distributions, which were mixture. From that procedure, several kinds of statistical distributions could be generated. Only one restriction was used: at least 30% of data generated by a statistical distribution should be intersection with another one. This makes a more realistic simulation, when compared with real data provide from physicians. For each of four classes were generated 30 thousand vectors, from which 10 thousand were used to calibrate the system and the other 20 thousand were used to check the assessment method based on GBN. Each vector contained 10000 positions, which is equivalent to 10 seconds simulation and the calibration time to build the General Bayesian Network was 19 minutes and 6 seconds.

A comparison of the classification agreement between SUAS based on GBN and the generated data was performed using the Cohen's Kappa Coefficient, according to recommended by literature [4] because it is known to be over conservative. The classification matrix obtained is presented in the Table 1. Each line of that matrix represents the real class of performance which data belongs and the column represents class assign by the SUAS. The Kappa coefficient was K=99.9767% with variance 3.8882×10^{-9}. In 14 cases, the SUAS based on GBN made mistakes and all of them were made in classes of performance C1 and C2. That performance is good and shows that General Bayesian Network is a competitive approach in the solution of assessment problems.

Table 1. Assessment results using a SUAS based on General Bayesian Network

	C_1	C_2	C_3	C_4
C_1	19994	6	0	0
C_2	8	19992	0	0
C_3	0	0	20000	0
C_4	0	0	0	20000

Those tests were performed in a commodity computer (Dual Core computer with 256Mb graphic card and 2 Gb of RAM). That computer run Linux operational system and CyberMed version 1.2. The average of CPU time consumed for assessment of a sample vector contained 10000 elements was 0.5 ms, i.e. that system is able to assess an user in lower than one second and characterizes an online assessment.

5 Comparison with SUAS Based on CBR

Some comparison tests were made to check performance of SUAS based on GBN in relation to the SUAS based on CBR [9], which was implemented in the CyberMed. Tests were performed with respect to time of execution, time of parameters learning and accuracy of that assessment for each SUAS.

All comparative tests were made in the same commodity computer using the same methodology (10 thousand vectors were used to calibrate the system and the other 20 thousand vectors were used to check the performance of the SUAS based on CBR), where each vector contained 10000 elements. The calibration time to build the Classical Bayes Rule was 7 minutes and 50 seconds.

The classification matrix obtained for the CBR is presented in the Table 2. The Kappa coefficient was K=92.7267% with variance 1.1403×10^{-6}. Besides the SUAS based on CBR made correct all assessments for classes of performance C3 and C4, it made several mistakes for C1 and C2 class. The reason for that bad performance can be credited to the restriction made in simulation, which at least 30% of data generated by a statistical distribution should be intersection with another one. That number of assessment errors can be considered high in a condition of medical training assessment, where the patient's life may depend on the physician skills.

Performance tests were conducted for both SUAS, by varying the number of elements in vector, between 1000 and 20000 elements, adding 1000 elements in vector to each test. The Figure 2 presents a comparative graph of performance for two SUAS, in which is possible to note the superior computational performance of SUAS based on CBR over SUAS based on GBN for all number of elements in vector, which were analyzed.

From the Figure 2, it is possible to infer when the performance of the SUAS based on GBN will not be sufficient to be characterized as online: around 40000 elements in the vector for that commodity computer used. However, for those same situations, the SUAS based on CBR probably reach that limit in around 100000 elements in vector.

Table 2. Assessment results using a SUAS based on Classical Bayes Rule

	C_1	C_2	C_3	C_4
C_1	16208	3792	0	0
C_2	0	19428	572	0
C_3	0	0	20000	0
C_4	0	0	0	20000

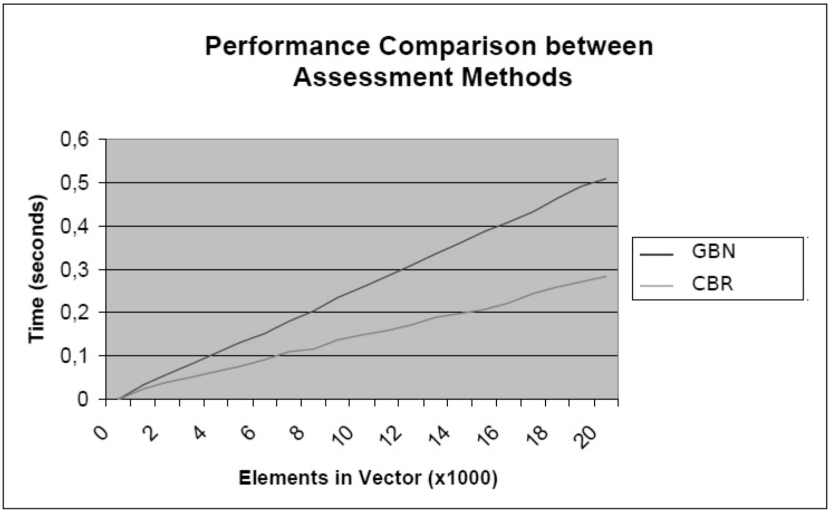

Fig. 2. Performance comparison between SUAS based on GBN and based on CBR according to number of elements in vector

6 Conclusions

In this paper, it was proposed a SUAS based on GBN, which was implemented in CyberMed and is available in the last version of that framework. Tests were performed using statistical simulation and results showed that assessment methodology as an appropriate approach in the solution of assessment problems, with high accuracy and low computational complexity. A comparison of this approach with another SUAS based on CBR was performed.

Quantitative data was generated according to the medical procedure chosen as case for the SUAS. This data generated by Gaussian distribution guarantee that both SUAS used the same data in its calibration and in the comparison of its performance. Both SUAS are able to perform online assessments and could be applied in other kind of training. However, the SUAS based on CBR is faster and can support around 100000 elements in vector, while the SUAS based on GBN can lead with 40000 elements approximately. It is important to note that in all those tests a commodity computer was used. The SUAS based on GBN achieve better results according to Cohen's Kappa Coefficient and for any number of elements in vector its accuracy is always higher from the same input vectors.

A performance statistical comparison with others SUAS, as well as with other kind of computer configurations, are planning as future works to achieve application limits for SUAS based on GBN.

Acknowledgments. This work is partially supported by Brazilian Council for Scientific and Technological Development, CNPq (Processes 310339/2009-0, 312375/2009-3 and 573710/2008-2).

References

1. Burdea, G., Coiffet, P.: Virtual Reality Technology, 2nd ed. Wiley Interscience (2003)
2. Cheng, J., Greiner, R.: Learning Bayesian Belief Network Classifiers: Algorithms and System. In: Proc. of the Fourteenth Canadian Conference on Artificial Intelligence (2001)
3. Dinsmore, M.: VR simulation: training for palpation of subsurface tumors. Master's Thesis, Dep. Mechanical and Aerospace Eng., Rutgers University (October 1996)
4. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. Wiley-Interscience (2000)
5. Gaba, D.M., et al.: Assessment of clinical performance during simulated crises using both technical and behavioral ratings. Anesthesiology 89, 8–18 (1998)
6. Krause, P.J.: Learning Probabilistic Networks. Knowledge Engineering Review 13, 321–351 (1998)
7. Machado, L.S., et al.: A Virtual Reality Simulator for Bone Marrow Harvest for Pediatric Transplant. Studies in Health Technology and Informatics 81, 293–297 (2001)
8. Machado, L.S., et al.: A Framework for Development of Virtual Reality-Based Training Simulators. Studies in Health Technology and Informatics 142, 174–176 (2009)
9. Machado, L.S., Moraes, R.M.: Intelligent Decision Making in Training Based on VR. In: Ruan, D. (org.) Computational Intelligence in Complex Decision Systems, ch. 4, pp. 85–123. Atlantis Press (2010)
10. Moraes, R.M., Machado, L.S.: Fuzzy Gaussian Mixture Models for On-line Training Evaluation in Virtual Reality Simulators. In: Int. Conference on Fuzzy Information Processing, China, pp. 733–740 (2003)
11. Moraes, R.M., Machado, L.S., Souza, L.C.: Online Assessment of Training in Virtual Reality Simulators Based on General Bayesian Networks. In: VI International Conference on Engineering and Computer Education (CDROM) (2009)
12. Moraes, R., Machado, L.: Development of a Medical Training System with Integration of Users' Assessment. In: Kim, J.-J. (ed.) Virtual Reality, pp. 325–348. Intech, Vienna (2011)
13. Moraes, R.M., Rocha, A.V., Machado, L.S.: Intelligent Assessment Based on Beta Regression for Realistic Training in Medical Simulators. Knowledge-Based Systems (in press, 2012)
14. Neapolitan, R.E.: Learning Bayesian Networks. Prentice Hall Series in Artificial Intelligence (2003)
15. Scott, et al.: Measuring Operative Performance after Laparoscopic Skills Training: Edited Videotape versus Direct Observation. Journal of Laparoendoscopic & Advanced Surgical Techniques 10(4), 183–190 (2000)
16. Wilson, M.S., et al.: MIST VR: a virtual reality trainer for surgery assesses performance. Annals of the Royal College of Surgeons of England 79, 4034 (1997)