Assessment Systems for Training Based on Virtual Reality: A Comparison Study

Ronei M. Moraes and Liliane S. Machado

Laboratory of Technologies for Virtual Teaching and Statistics (LabTEVE)
Federal University of Paraíba (UFPB)
João Pessoa, Brazil
ronei@de.ufpb.br, liliane@di.ufpb.br

Abstract—Training systems based on virtual reality have been used in several areas. In these systems users are immersed into a virtual and interactive environment to perform realistic training. In this paper, we present some of the challenges to construct a medical simulator based on virtual reality. Among them, the assessment allows to know users' performance in the training to analyze if they are prepared to perform the procedure in real situations. In order to choose an appropriate assessment method, this paper brings a comparison among four methods for training assessment.

Keywords—medical simulation, pattern recognition, virtual realty, fuzzy sets.

I. INTRODUCTION

Training systems based on virtual reality (VR) have been used in several areas [4]. In those systems, the user is immersed into a virtual world to perform realistic training with realistic interactions. However, it is important to assess users´ training to know the quality of their skills. Several kinds of training based on VR use to record the user actions in videotapes to post-analysis by experts [3]. In these cases, user receives their assessment after some time (days or weeks). This is a problem because probably after some hours the user will not remember their exact actions, which will make difficult the use of the assessment information to improve their performance. Additionally, several kinds of training cannot be simply classified as bad or good due to their complexity. Then, the existence of an online assessment tool incorporated into a simulation system based on virtual reality is important to allow the learning improvement and users assessment [21]. An online assessment system allows the user to improve their learning because it can identify, immediately after the training (feedback in less than 1 second), where mistakes occurred or actions presented low efficiency.

The research area on training assessment for simulators based on VR is only 15 years old [4]. The early works in that area were probably proposed by Dinsmore et al. [9,10,20] that used a quiz to assess users of a VR environment to identify subcutaneous tumors. The quiz contained questions related to the diagnosis and hardness of tumor. Similarly, Wilson et al. [51] created a minimally invasive system (MIST) 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 statistical methods.

In parallel, methods to assess surgical skills have been developed by several research groups. Some of them use statistical models to do that offline [8] and others use statistical methods to show that through VR based systems it is possible to discriminate between expert and novice physicians [14,49]. It has been shown also that surgeons trained in VR systems could obtain better results [13] when compared to others trained by traditional methods. Additionally, the assessment of psychomotor skills in VR systems that include haptic devices can quantify surgical dexterity with objective metrics [6]. Thus, VR systems for training can be used to provide metrics to a proficiency criterion of learning [6,15]. Due to those reasons, McCloy and Stone [28] pointed out the assessment of psychomotor skills as the future of medical teaching and training.

The first proposal for online training assessment in VR systems was presented by Burdea et al. [3] and was based on a boolean logic that compared diagnoses provided by users with correct ones stored in the simulator. However, ordinary computers of that generation were not able to simultaneously run virtual reality environments/simulators and online assessment systems if several interaction variables were monitored. After that, more sophisticated assessment methods were proposed.

Because VR simulators are real-time systems, an assessment tool must continuously monitor all user interactions and compare their performance with pre-defined expert's classes of performance. Some benefits of online assessment are: a) user can quickly identify their mistakes and try to correct them in the next training session, b) the assessment can be used by the simulator to increase or decrease dynamically the level of difficulty.

In medicine, some models for offline [19,42,43] and online [12,16,21,29] assessment of training have been proposed. The main problems related to online training assessment methodologies applied to VR systems are the computational complexity and the accuracy. An online assessment tool must have low complexity in order to do not compromise VR simulations performance, but it also must have high accuracy in order to do not compromise the user assessment.

Some of the models previously mentioned are based on machine learning and use discretization of continuous variables, as proposed in [34] or some change of the problem

domain using fuzzy sets [36], based on Naive Bayes method or some modification of that. In this paper, we present some challenges related to the development of a medical simulator. Among them, the assessment allows to know users' performance in the training to analyze if they are prepared to perform the procedure in real situations. In order to choose an appropriate assessment method, this paper brings a comparison among four methods for training assessment.

The present paper is organized as follows: in Section 2 we introduce training systems based on VR. In Section 3, we give a historical review over assessment methods. In Section 4, we provide details about a bone marrow harvest simulator, which was used to compare the four methods. Theoretical aspects are presented in Section 5. In Section 6 we presented calibration aspects for all methods in order to allow performance comparisons. The results are discussed in Section 7. Finally, we present some concluding remarks in Section 8.

II.   TRAINING BASED ON VIRTUAL REALITY

Virtual Reality refers to real-time systems modeled by computer graphics that allow user interaction and movements with three or more degrees of freedom [4]. More than a technology, VR became a new science that joins several fields as computing, robotics, graphics, engineering and cognition, among others. VR worlds are 3D environments, created by computer graphics techniques, where one or more users are immersed totally or partially to interact with virtual elements. The realism of a VR application is given by the graphics resolution and by the exploration of users senses. Mainly, special devices stimulate the sight, hearing and touch. As example, head-mounted displays (HMD), or even ordinary monitors combined with special glasses, can provide stereoscopic visualization, multiple sound sources positioned provide 3D sound, and touch can be simulated through haptic devices [2].

VR systems for training can provide significant benefits over other training methods since users can perform procedures in a safe way as many times as necessary to understand or acquire skills. Additionally, materials do not wear out and users' interactions can be monitored. These features are particularly interesting in training of dangerous activities.

In some cases, the procedures are performed without any kind of visualization and the only information received is noticed by the touch sensations provided by robotic devices with touch and force feedback, called haptic devices. These devices can measure forces and torque applied during the user interaction [2] and the data can be used in an assessment [21,42]. A specific kind of haptic device, as the presented in Figure 1, is based on a robotic arm and provides force feedback and tactile sensations during user manipulation of objects in a three dimensional scene. This way, user can feel objects texture, density, elasticity and consistency. Since the objects have physical properties, the user can identify them in a 3D scene (without seeing them) by the use of this kind of device [41]. Haptic devices based on robotic arms are particularly interesting for medical applications due to their manipulation similarity when compared to real surgical tools.

According to Riva [40], medicine is an important and potential area for virtual reality applications since they can provide safe, repetitive and diversified training. Despite the advances, the technology available still do not allow the creation of a realistic simulator to explore the whole human body using volumetric models, 3D visualization, multiple haptics, deformation, interactive cut and assessment. The use of all these features requires a lot of processing and can compromise real-time performance. In haptics, for example, it is commonly used a single point of contact for touch and force feedback due to the lower processing and device costs if compared to devices that provide multiple contact point. Then, VR simulators for medicine are defined and developed to deal with specific parts of the human body and present some limitations: according to the procedure requirements, some VR techniques are more explored than others. These limitations do not impede the development of good and useful training simulators [4]. However, in spite of the several developments for laparoscopy [4], breast surgery [1], stroke rehabilitation [5], heart surgery [46] and hepatic surgery [7], among others, assessment approaches of the training performed by users is little explored.

Figure 1.   Phantom Omni haptic device used in VR systems and in the Bone Marrow Harvest simulator.

In order to decrease the development time of simulators, several frameworks were proposed. Those software packages intend to provide a set of functionalities that can be easily reused to compose training and visualizing applications for medical purposes [37]. CyberMed is one among those frameworks that presents online assessment methods that can be used in the development of training simulators. The design of CyberMed included the use of several design patterns that allow the framework presenting interfaces to add new methods and devices support [37].

The several functionalities of a framework demand previous decision of which approach will need to be used. Then, the designer must choose one of the available visualization methods. It also happens to the assessment method that must be chosen according to the simulator features.

III.   ASSESSMENT IN VIRTUAL REALITY SIMULATORS

The assessment in simulations is necessary to monitor the training and provide feedback about the user performance. User movements, as spatial movements, can be collected from mouse, keyboard and any other tracking device. Applied forces, angles, position and torque can be collected from haptic

devices. Then, virtual reality systems can use one or more variables, as the mentioned above, to assess a simulation performed by user.

Some simulators for training present a method of assessment. However they just compare the final result with an expected result or are post-analyses of videotape records [3]. Models for offline or online assessment of training have been proposed, some of them use Discrete Hidden Markov Models [25,26,42,44,45] or Continuous Hidden Markov Models [30,43] to modeling forces and torque during a simulated training in a guinea pig. Machado et al. [21] proposed the use of a fuzzy rule-based system to online assessment of training in virtual worlds. Using an optoelectronic motion analysis and video records, McBeth et al. [27] acquired and compared postural and movement data of experts and residents in different contexts by use of distributions statistics. In a previous work we proposed several methods for online assessment [29,31,32,34,35]. We also proposed a methodology to automatically assess a user's progress to improve their performance in virtual reality training systems [33] using statistical measures and models (time dependent or not) as well as a fuzzy expert system. After that, Morris et al. [38] suggested the use of statistical linear regression to evaluate user's progress in a bone surgery.

Although various methods of training assessment based on virtual reality can be found in the literature, it is important to highlight that the choice of the method of assessment depends on the kind of training, particularly on variables that can be measured during the training execution. The variability and the statistical measures of these variables characterize the users´ skills and classes of performance can be assigned for them. An assessment method must provide high accuracy in order to not compromise the user assessment and as a computational tool for online training assessment must have low complexity in order to do not compromise VR simulations performance.

In this paper, we present four systems for training assessment based on VR, whose theoretical aspects are presented in Section 5. These systems can perform an online training assessment for virtual reality simulators. They use a vector of information, with data collected from user interactions with virtual reality simulator, and these data are compared by the assessment system with M pre-defined classes of performance. The assessment tools were developed and analyzed for online evaluation of users of a bone marrow harvest simulator.

IV. BONE MARROW HARVEST SIMULATOR BASED ON VIRTUAL REALITY

The bone marrow harvest is one of the stages of the bone marrow transplant and demands dexterity from the physician who performs it. Basically, two steps compose it: palpation and harvest [39]. In the palpation step, the physician must identify the iliac crest under the skin of the pelvic region. The iliac crest is the region used to perform the second step: the harvest of bone marrow.

A multidisciplinary team must compose the development of a simulator for medical training [24]. It includes the presence of a physician or expert in the procedure. For the bone marrow harvest, a physician provided all details about the procedure, approved the approach, calibrated and validated the system [22]. Thus, the simulator developed presents the two steps of the harvest procedure. Since the simulator is concerned with training, an extra step was added to allow studying the anatomy of the pelvic region. Thus, three modules compose the final application: Study, Palpation and Harvest. These modules were implemented to have the same modules and functions described in a previous work [22], but used the framework CyberMed [37]. This framework allows quick integration of the several tasks that compose a simulator since it can synchronize and optimize them to guarantee consistent execution in real time. Souza et al. [47] presented a comparison related to the previous work and this new implementation.

The bone marrow harvest simulator used the CyberMed classes to provide the functionalities of visualization, storage and management of 3D models, interaction control, haptic control and online user's assessment. The assessment class, called CybAssess, is available in the CyberMed and can be used to collect user actions during the simulation [23]. It also provides a default interface to integrate assessment methods, as those presented in this paper. The models of human body structure used were the same of a previous work [22] and represent the skin of the pelvic region, the iliac bone and the bone marrow. Figure 2 shows the models used to represent the interaction object in the visualization. The CybHaptics [48] allowed relating the contact point to a vertex of the objects: a point in the tip of finger and tip of the needle.
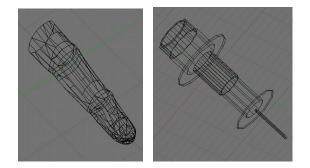

Figure 2. Models used to represent the interaction object in the visual scene: a finger and a needle.

A menu was designed to offer to the user three modules for training. This menu is dynamically modified according to the user choices. Only visual exploration is enabled in the Study module and the user can choose which structures they want to see and set their transparency. Figure 3 shows the Study module and the menu options available. The user can also modify the position and orientation of the structures through mouse interaction. If shutter glasses are available, it is possible to start the stereoscopy visualization.

The second module available is the Palpation module. For this step of the training the position of the objects was fixed and the transparent view option was disabled. In fact, the only visualization possible is of the skin model. Because the visualization of the bone and bone marrow models is not allowed, these models were disabled and will not be rendered.

However, they can be identified by touch when the user starts the haptic interaction. With the haptic device the user will be able to feel the different material properties throughout the skin and identify a harder area, located under the iliac crest (not visible). The approach used was based on meshes. Thus, to allow the identification of the iliac crest, two small spheres were positioned over the iliac crest and received the same haptic properties of the bone. These spheres cannot be seen and are located a little off the skin. Thus, the user can experiment a different hardness when touch the region. This approach did not compromise the realism and has as goal just to allow users the identification of the correct place to insert the needle. The haptic device cannot penetrate in the model and only the touch is allowed.
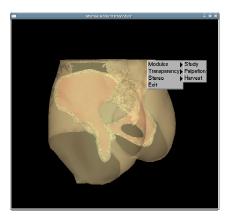

Figure 3.   Screenshot of the Study module and the menu options available.

In this module, the finger model was related to the haptic device and the point of contact is located on its tip. Due to technological limitations, only one contact point is available, since the haptic device used (Figure 1) cannot deal with multiple contact points. However, the movements performed by user using the haptic device when touching the virtual model allow the identification of the tissue properties.

In the last module, the Harvest, the user can practice how to harvest the bone marrow. As in the Palpation module, movements with the body models are not allowed. Now, a needle represents the haptic device. In this module it is possible to penetrate into the models with the haptic device and all body structures - skin, bone and bone marrow - are haptically displayed. It allows reaching the bone marrow under the skin and inside the bone.

To calibrate the assessment tool, the simulation had to be executed several times by an expert. The expert executes the procedure several times and labels each one according to one of the M classes available. An expert must do this task since he knows the particularities and weight of mistakes that a user can make. This stage allows acquiring the assessment parameters to be used by the online assessment method. A specialist in bone marrow harvest carried it out.

V.   OVERVIEW OF PROBABILISTIC METHODS SUITABLE FOR ASSESSMENT

In this section, we present a brief comparison among four statistical methods for training assessment identified as adequate for the bone marrow harvest simulator. An advantage of these assessment methods is to allow inclusion of other variables in the assessment tool with low performance degradation of the virtual reality simulation.

A.   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 a user. We can determine the most probable class of a vector of training data $\mathbf{X}$, according to sample data D, where $\mathbf{X}$ is a vector with n features obtained when a training is performed, i.e. $\mathbf{X}=\{X_1, X_2, ..., X_n\}$. Using the Bayes Theorem [17]:

$$P(w_i \mid \mathbf{X}) = [P(\mathbf{X} \mid w_i)\, P(w_i)] / P(\mathbf{X}), \text{ where } i \in \Omega \quad (1)$$

The classification rule is performed according to

$$\mathbf{X} \in w_i \text{ if } P(w_i \mid \mathbf{X}) > P(w_j \mid \mathbf{X}) \text{ for all } i \neq j \text{ and } i, j \in \Omega \quad (2)$$

As $P(X)$ is the same for all classes w_i, then it is not relevant for data classification. In Bayesian theory, $P(w_i)$ is called a priori probability for w_i and $P(w_i \mid X)$ is a posteriori probability for w_i where $\mathbf{X}$ is known. Then, the classification rule done by eq. 2 is modified:

$$\mathbf{X} \in w_i \text{ if } P(\mathbf{X} \mid w_i)\, P(w_i) > P(\mathbf{X} \mid w_i)\, P(w_i)$$

$$\text{for all } i \neq j \text{ and } i, j \in \Omega \quad (3)$$

The eq. 3 is known as Bayesian decision rule of classification. However, in the cases, which $\mathbf{X}$ can assume statistical Gaussian distribution, it can be convenient to use [17]:

$$g(\mathbf{X}) = \ln [P(\mathbf{X} \mid w_i)\, P(w_i)]$$

$$= \ln [P(\mathbf{X} \mid w_i)] + \ln[P(w_i)], i \in \Omega \quad (4)$$

where $g(\mathbf{X})$ is known as discriminant function. We can use eq. 4 to modify the formulation done by Bayesian decision rule in eq. 3:

$$\mathbf{X} \in w_i \text{ if } g_i(\mathbf{X}) > g_j(\mathbf{X}) \text{   for all } i \neq j \text{ and } i, j \in \Omega \quad (5)$$

It is important to note that if statistical distribution of training data can assume multivariate Gaussian distribution, the use of eq. 5 has interesting computational properties. As example, mathematical simplifications are possible and they provide lower computational complexity [17].

B.   Fuzzy Sets

In classical set theory a set A of a universe X can be defined by a membership function $\mu_A(x)$, with $\mu_A: X \rightarrow \{0,1\}$, where 1 means that x is included in A and 0 means that x is not included in A. A fuzzy set can be seen as a representation in classical set theory, of which we only have an imperfect knowledge. In this case, the membership function cannot be done by only one value 0 or 1, but by a value in $[0,1]$ interval [53].

The probability of a fuzzy event is defined by Zadeh [54]: let be (R^n, ϕ, P) a space of probability where ϕ is an σ-algebra in R^n and P is a probability measure over R^n. Then a fuzzy event in R^n is a set A in R^n, with membership function $\mu_A(x)$, where $\mu_A: R^n \rightarrow \{0,1\}$ is Borel-mensurable. The probability of a fuzzy event A is defined by Lebesgue-Stieltjes integral:

$$P(A) = \int_{R^n} \mu_A(x)\, dP = E(\mu_A) \qquad (6)$$

In other words, the probability of a fuzzy event A with membership function μ_A is the expected value of the membership function μ_A.

C. Fuzzy Bayes Rule

Again, let the classes of performance for a user done by w_i, $i=1,...,M$, where M is the total number of classes of performance. However, now we assume that w_i are fuzzy sets over space of decision Ω. Let be $\mu_{wi}(X)$ the fuzzy membership function for each class w_i given by a fuzzy information source (for example, a rule composition system of the expert system, or a histogram of the sample data), according to a vector of data X. In our case, we assume that fuzzy information font is a histogram of the sample data.

By use of fuzzy probabilities and fuzzy Bayes rule [50] in the classical Bayes rule [11], we have the fuzzy probability of the w_i class, given the vector of data X:

$$P(w_i \mid X) = [\mu_{wi}(X)\, P(X \mid w_i)\, P(w_i)] / P(X),$$

$$\text{with } \sum_i \mu_{wi}(X)=1, i \in \Omega \qquad (7)$$

However, as the denominator is independent, then the Fuzzy Bayes classification rule is to assign the vector of training data X from the user to w_i class of performance if:

$$X \in w_i \text{ if } \mu_{wi}(X)\, P(X \mid w_i)\, P(w_i)$$

$$= \max j \leq M\ \mu_{wj}(X)\, P(X \mid w_j)\, P(w_j) \qquad (8)$$

D. Naive Bayes Method

A Naive Bayes classifier computes conditional class probabilities and then predicts the most probable class of a vector of training data $X=\{X_1, X_2, ..., X_n\}$, according to sample data D. From eq. 1:

$$P(w_i \mid X) = [P(X \mid w_i)\, P(w_i)] / P(X) \Leftrightarrow P(w_i \mid X_1, X_2, ..., X_n) =$$

$$= [P(X_1, X_2, ..., X_n \setminus w_i)\, P(w_i)] / P(X) \qquad (9)$$

However, as $P(X)$ is the same for all classes w_i, then it is not relevant for data classification and can be rewritten as:

$$P(X \mid w_i)\, P(w_i) = P(X_1, X_2, ..., X_n \setminus w_i)\, P(w_i) \qquad (10)$$

The eq. 10 is equivalent to the joint probability model:

$$P(X_1, X_2, ..., X_n \setminus w_i)\, P(w_i) = P(X_1, X_2, ..., X_n, w_i) \qquad (11)$$

Now, using successive applications of the conditional probability definition over eq. 11, the following can be obtained:

$$P(X_1, X_2, ..., X_n, w_i) = P(w_i)\, P(X_1, X_2, ..., X_n \setminus w_i) =$$

$$= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2, ..., X_n \setminus w_i, X_1)$$

$$= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i, X_1)\, P(X_3, ..., X_n \setminus w_i, X_1, X_2)$$

$$...$$

$$= P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i, X_1 \,...P(X_n \setminus w_i, X_1, X_2,...,X_{n-1})$$

The Naive Bayes classifier receives this name because its naive assumption of each feature X_k is conditionally independent of every other feature X_l, for all $k \neq l \leq n$. It means that knowing the class is enough to determine the probability of a value X_k. This assumption simplifies the equation above, due to:

$$P(X_k \setminus w_i, X_l) = P(X_k \setminus w_i) \qquad (12)$$

for each X_k and the eq. 11 can be rewritten as:

$$P(X_1, X_2, ..., X_n, w_i) =$$

$$P(w_i)\, P(X_1 \setminus w_i)\, P(X_2 \setminus w_i).\, P(X_n \setminus w_i) \qquad (13)$$

unless a scale factor S, which depends on $X_1, X_2, ..., X_n$. Finally, eq. 11 can be expressed by:

$$P(w_i \mid X_1, X_2, ..., X_n) = (1/S)\, P(w_i)\, \Pi^n_{k=1}\ P(X_k \setminus w_i) \qquad (14)$$

Then, the classification rule for Naive Bayes is done by:

$$X \in w_i \text{ if } P(w_i \mid X_1, X_2, ..., X_n) > P(w_j \mid X_1, X_2, ..., X_n)$$

$$\text{for all } i \neq j \text{ and } i, j \in \Omega \qquad (15)$$

and $P(w_* \mid X_1, X_2, ..., X_n)$ with $* = \{i, j \mid i, j \in \Omega\}$, is done by eq. 14.

To estimate parameters for $P(X_k \setminus w_i)$ for each class i, a maximum likelihood estimator, named P_e, is used:

$$P_e(X_k \setminus w_i) = \#(X_k, w_i) / \#(w_i) \qquad (16)$$

where $\#(X_k, w_i)$ is the number of sample cases belonging to class w_i in all sample data D and having the value X_k, $\#(w_i)$ is the number of sample cases that belong to the class w_i in all sample data D.

E. Gaussian Naive Bayes (GNB)

As mentioned above, the NB Method must be applied over discrete or multinomial variables. Some approaches were developed to use NB Method with continuous variables, as several discretization methods [18,52] were used in the first stage to allow the use of the Naive Bayes method later. However, this approach can affect classification bias and variance of the NB method. Other approach is use Gaussian distribution for X and to compute its parameters from D, i.e., mean vector and covariance matrix [17]. From eq. 14 and using some mathematical simplification, it is possible to reduce computational complexity of that equation:

$$\log [P(w_i \mid X_1, X_2, ..., X_n)]$$

$$= \log [(1/S)\, P(w_i)\, \Pi^n_{k=1}\ P(X_k \setminus w_i)]$$

$$= \log (1/S) + \log P(w_i) + \sum^n_{k=1}\ \log[P(X_k \setminus w_i)] \qquad (17)$$

As S is a scale factor, it is not necessary to be computed in classification rule for GNB. Then:

$$X \in w_i \text{ if}$$

$$\{\log P(w_i) + \sum\nolimits_{k=1}^{n} \log[P(X_k \setminus w_i)]\} >$$

$$\{\log P(w_j) + \sum\nolimits_{k=1}^{n} \log[P(X_k \setminus w_j)]\}$$

$$\text{for all } i \neq j \text{ and } i, j \in \Omega \qquad (18)$$

Based on the same space of decision with M classes, a GNB method computes conditional class probabilities and then predicts the most probable class of a vector of training data $\mathbf{X}$, according to sample data D. The parameters of GNB method are learning from data and the conditional probabilities are estimated using eq. 4 and the final decision about vector of training data $\mathbf{X}$ is done by eq. 18.

VI.   CALIBRATION OF THE ASSESSMENT TOOL

An assessment tool should supervise the user's movements and other parameters associated to them. The system must collect information about positions in space, forces, torque, resistance, speeds, accelerations, temperatures, visualization position and/or visualization angle, sounds, smells and etc. The virtual reality simulator and the assessment tool are independent systems, however they act simultaneously. The user's interactions with the simulator are monitored and the information is sent to the assessment tool that analyzes the data and emits a report on the user's performance at the end of the training. Depending on the application, all those variables or some of them will be monitored (according to their relevance to the training).

The virtual reality system used for all tests is the bone marrow harvest simulator [22]. In a first movement on the real procedure, the trainee must feel the skin of the human pelvic area to find the best place to insert the needle used for the harvest. After, they must feel the tissue layers (epidermis, dermis, subcutaneous, periosteum and compact bone) trespassed by the needle and stop at the correct position to do the bone marrow extraction. In our VR simulator the trainee interacts with a robotic arm and their movements are monitored in the system through some variables [22]. For reasons of general performance of the VR simulator, the following variables were chosen for monitoring: spatial position, velocities, forces and time on each layer. Previously, an expert, according to M classes of performance defined by him, calibrated the system. The calibration process consists in to execute several times the procedure and to classify each one according to classes of performance. The number of classes of performance was defined by an expert as M=3: 1) correct procedures, in which procedure is performed perfectly; 2) acceptable procedures, in which minor mistakes were done, but they are mistakes that did not compromise the execution of the procedure neither patient's health, 3) badly executed procedures, in which mistakes are serious and could compromise the execution of the procedure or patient's health. So, the classes of performance for a trainee could be: "you are well qualified", "you need some training yet", and "you need more training".

The information of variability about these procedures is acquired using the methods presented in previous sections. In our case, we assume that the source of information for w_i classes is the vector of the sample data D. The user makes their training in the virtual reality simulator and the assessment tool collects data from their manipulation. All probabilities of data for each class of performance are calculated and at the end the user is assigned to a w_i class of performance. So, when a trainee uses the system, their performance is compared with each expert's class of performance and the assessment tool assigns the most appropriated class, according to the trainee's performance. At the end of the training, the assessment system reports the classification to the trainee.

The Cohen's Kappa Coefficient was used to perform the comparison of the classification agreement between expert and each assessment tool. There are other methods, but Kappa Coefficient is known to be over conservative, as recommended in the pattern recognition and classification literatures [11].

To perform all simulations, it was used the same computational platform. A Pentium IV PC compatible, 2GB of RAM and 80GB of hard disk composed it.

VII.   COMPARISON EXPERIMENTAL STUDY

In this section, a short comparison among assessment methods used in the bone marrow harvest simulator is presented. To perform those comparisons, all assessment tools were configured and calibrated by an expert for the same three classes used before. The same sixty samples of training (twenty of each class of performance) were used for calibration of all assessment systems. Analogously, the data of the same 150 procedures from users training were used for a controlled and impartial comparison among the assessment systems.

A. Classical Bayes Rule

The classification matrix obtained for the Assessment Tool based on Classical Bayes Rule (ATBCBR) is presented in the Table 1. The Kappa coefficient was K=81.0% with variance 1.6×10^{-3} %. In 19 cases, the evaluation tool made mistakes and at least one classification was made incorrectly in all classes. That performance is good and shows that an ATBCBR is a competitive approach in the solution of assessment problems.

TABLE I.   CLASSIFICATION MATRIX FOR THE ASSESSMENT TOOL BASED ON CLASSICAL BAYES RULE

Class of performance according to experts	Class of performance according to Evaluation Tool based on Classical Bayes Rule		
	1	2	3
1	46	0	4
2	1	49	0
3	8	6	36

When the ATBCBR performed the classification, few mistakes were observed. However, it is possible to see by Tables 1 and 2 and by Kappa coefficients that the performance of the classification based on Classical Bayes Rule is lower than the one based on Fuzzy Gaussian Naive Bayes. In statistical terms, the difference of performance between those assessment methods is significant. About computational performance the average of CPU time consumed for ATBCBR was 0.0160 seconds.

B. Fuzzy Bayes Rule

These same cases were labeled using the Assessment Tool based on Fuzzy Bayes Rule (ATBFBR) and generated the classification matrix showed in Table 2. The diagonal of that matrix shows the correct classification. In the other cells, we can observe the mistakes of classifications.

From the classification matrix obtained, the Kappa coefficient for all samples was K=90.0% with variance 9.3×10^{-4} %. In only 10 cases, the evaluation tool made mistakes. It is important to notice that for the class "acceptable procedures", all classifications were correct. That performance is very acceptable and shows the good adaptation of ATBFBR in the solution of this evaluation problem.

TABLE II.   CLASSIFICATION MATRIX FOR THE ASSESSMENT TOOL BASED ON FUZZY BAYES RULE

Class of performance according to experts	Class of performance according to Evaluation Tool based on Fuzzy Bayes Rule		
	1	2	3
1	44	0	6
2	0	50	0
3	1	3	46

Another important result is the computational performance of the evaluation tool: with the same Pentium IV PC compatible, the average time of CPU consumed by the evaluation was 0.0310 seconds of CPU. Then, we can affirm that the ATBFBR has low computational complexity.

C. Naive Bayes

The classification matrix obtained for the Assessment Tool based on Naive Bayes (ATBNB) is presented in the Table 3. The Kappa coefficient was K=66.0% with variance 2.7×10^{-3} %. In 34 cases, the assessment tool made mistakes. However, that performance is acceptable and shows that a ATBNB is a competitive approach in the solution of assessment problems. About computational performance of the Assessment Tool, the one based on FGNB was faster than the one based on NB. The average of CPU time consumed for assessment of training based on NB was 0.0940 seconds of CPU using the same Pentium IV PC compatible.

TABLE III.   CLASSIFICATION MATRIX FOR THE ASSESSMENT TOOL BASED ON NAIVE BAYES

Class of performance according to experts	Class of performance according to Evaluation Tool based on Naive Bayes		
	1	2	3
1	36	1	13
2	0	41	9
3	9	2	39

D. Gaussian Naive Bayes

These same cases were labeled using the Assessment Tool Based on Gaussian Naive Bayes (ATBGNB) and it generated the classification matrix showed in Table 4. From that classification matrix, the Kappa coefficient for all samples was K=80.00% with variance 1.7×10^{-3} %. In only 20 cases, the assessment tool made mistakes. It is important to note that for the class "acceptable procedures", only one classification was incorrect. That performance is very acceptable and it shows the good adaptation of ATBGNB in the solution of this assessment problem.

Another important result is the computational performance of the assessment tool: with the same Pentium IV PC compatible, the average time of CPU consumed by the assessment was 0.0150 seconds of CPU. Then, we can affirm that the ATBGNB has low computational complexity.

TABLE IV.   CLASSIFICATION MATRIX FOR THE ASSESSMENT TOOL BASED ON GAUSSIAN NAIVE BAYES

Class of performance according to experts	Class of performance according to Evaluation Tool based on Gaussian Naive Bayes		
	1	2	3
1	45	0	5
2	1	49	0
3	8	6	36

It is important to note that there are significant statistical differences among all of them and the best results were provide by Assessment Tool based on Fuzzy Bayes Rule (Kappa coefficient equal 90.0%). However, the faster assessment is provided by Assessment Tool Based on Gaussian Naive Bayes which consumed 0.0150 seconds of CPU. The Assessment Tool based on Classical Bayes Rule is a competitive and fast approach when all statistical distributions of data can be considered Gaussian distributions. Otherwise, an Assessment Tool based on Naive Bayes is more adequate. This last assessment tool showed the slowest among the methods in this comparison. However, it was observed that all approaches can be considered competitive taking into account its statistical particularities.

The Figure 4 summarizes accuracy of each method, measured by Cohen´s Kappa Coefficient (at left side) and computational performance for each method using CPU time (right side). Note that the Assessment Tool based on Naive Bayes presented the lowest accuracy and also the largest computational time spent to perform the assessment. By other hand, the Assessment Tool based on Fuzzy Bayes Rule provided higher precision in results, but the Assessment Tool Based on Gaussian Naive Bayes provided the faster assessment.
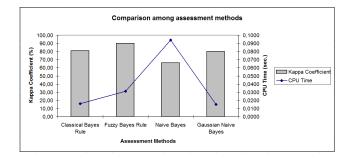
Figure 4.   Comparison chart among assessment methods.

VIII.   CONCLUSIONS AND FUTURE WORKS

The main challenge found to construct a medical simulator based on virtual reality is the integration of several tasks to real-time execution. It includes the choice of adequate methods for each task present in the simulator. However, a simulator for training must include an assessment module to allow users to know and improve their learning and skills. An important challenge in the use of online evaluation is collect and process users interaction without commit the performance of the application.

In this paper we presented a simulator for bone marrow harvest training based on virtual reality composed by several modules, as visualization, haptics interaction, collision detection and users assessment, among others. The presented simulator was used to compare four methodologies for online assessment found in the literature. In this case, the Assessment Tool based on Fuzzy Bayes Rule provided higher precision in results, but the Assessment Tool Based on Gaussian Naive Bayes provided the faster assessment. In general terms, the four methods used in this comparison are competitive approaches in the solution of assessment problems and are able to allow inclusion of other variables in the assessment tool with low degradation of the performance of the virtual reality simulation. Future works include the use of the assessment tools in other simulators, as well as the proposal and implementation of other methods for assessment.

The research related to online assessment has as goal to identify methods that could be used in real-time to reconfigure the simulation according to users' performance, besides the report their skills. In these cases, it would also be possible to use decision making embedded in the training applications to increase and decrease, automatically, the difficulty level. This paper intended to show some methodologies that can be used online in further works in this direction. However, the accuracy of assessment is also important to verify users skills. It is a theme of researches to allow certification of professionals in simulators.

ACKNOWLEDGMENT

REFERENCES

[1] R. Balaniuk, I. Costa, J. Melo, "Cosmetic Breast Surgery Simulation", Proc. Symp. on Virtual Reality, pp. 387-396, Belém, Brasil, 2006.

[2] G. Burdea, Force and Touch Feedback for Virtual Reality. Wiley-Interscience.1996.

[3] G. Burdea, G. Patounakis, V. Popescu, and R.E. Weiss, "Virtual Reality Training for the Diagnosis of Prostate Cancer". Proc. IEEE Virtual Reality Annual Int. Symposium (VRAIS'98), pp. 190-197, 1998.

[4] G. Burdea and P. Coiffet, Virtual Reality Technology. 2nd ed. Wiley Interscience, 2003.

[5] L. Cardoso, R. Costa, A. Piovesan, M. Costa, L. Penna, "Using Virtual Environments for Stroke Rehabilitation", Proc. IEEE 5th Int. Workshop on Virtual Rehabilitation, pp. 1-5, New York.

[6] A. Darzi, S. Smith, N. Taffinder, "Assessing operative skills", British Medical Journal, 318(7188):887-888, 1999.

[7] H. Dellingette, N. Ayache, "Hepatic Surgery Simulation", Communications of ACM, 48(2): 31-36. 2005.

[8] A.M. Derossis et al., "Development of a model for training and evaluation of laparoscopic skills", American Journal of Surgery 175(6): 482-487, 1998.

[9] M. Dinsmore, "Virtual reality simulation: training for palpation of subsurface tumors", Master's Thesis, Department of Mechanical and Aerospace Engineering, Rutgers University, October. 1996.

[10] M. Dinsmore, N. Langrana, G. Burdea, J. Ladeji, "Virtual Reality Training Simulation for Palpation of Subsurface Tumors", Proc. IEEE VRAIS 1997, pp. 54-60. 1997.

[11] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification. 2nd ed. Wiley, 2000.

[12] M. Färber et al., "Training and evaluation of lumbar punctures in a VR-environment using a 6DOF haptic device", Studies in Health Tech. and Informatics, 132: 112-114, IOSPress, 2008.

[13] A.G. Gallagher et al., "Virtual Reality training in laparoscopic surgery: A preliminary assessment of minimally invasive Surgical trainer Virtual Reality (MIST VR)", Endoscopy 31(4):310-313, 1999.

[14] A.G. Gallagher et al., "Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality", World Journal of Surgery 25(11):1478-1483, 2001.

[15] A.G. Gallagher et al., "Virtual Reality Simulation for the Operating Room: Proficiency-Based Training as a Paradigm Shift in Surgical Skills Training", Annals of Surgery 241(2): 364-372, 2005.

[16] J. Huang, S. Payandeh, P. Doris, and I. Hajshirmohammadi, "Fuzzy Classification: Towards Evaluating Performance on a Surgical Simulator", Studies in Health Tech. and Informatics, 111: 194-200, IOSPress, 2005.

[17] R.A. Johnson and D.W. Wichern, Applied Multivariate Statistical Analysis. Prentice Hall, 1998.

[18] I. Kononenko, "Inductive and Bayesian learning in medical diagnosis", Applied Artificial Intelligence, 7(4): 317–337, 1993.

[19] K. Kumagai, J. Yamashita, O. Morikawa, and K. Yokoyama, "A New Force-Based Objective Assessment of Technical Skills in Endoscopic Sinus Surgery". Studies in Health Tech. and Informatics 125: 235-237, IOSPress, 2007.

[20] N. Langrana, G. Burdea, J. Ladeji, M.Dinsmore, "Human Performance Using Virtual Reality Tumor Palpation Simulation", Computer & Graphics 21(4):451-458, 1997.

[21] L.S. Machado, R.M. Moraes, and M.K. Zuffo, "Fuzzy Rule-Based Evaluation for a Haptic and Stereo Simulator for Bone Marrow Harvest for Transplant". Proc. 5th Phantom Users Group Workshop (PUG'2000), 2000.

[22] L.S. Machado, A.N. Mello, R.D. Lopes, V. Odone Filho, and M.K. Zuffo, "A Virtual Reality Simulator for Bone Marrow Harvest for Pediatric Transplant". Studies in Health Technology and Informatics, 81: 293-297, IOSPress, 2001.

[23] L.S. Machado and R.M. Moraes, "Online Training Evaluation in Virtual Reality Simulators Using Evolving Fuzzy Neural Networks", Proc. 6th Int. FLINS Conf., pp. 314-317, 2004.

[24] L.S. Machado, R.M. Moraes, R.M., Intelligent Decision Making in Training Based on Virtual Reality. Book Chapter. Da Ruan. (Org.) Computational Intelligence in Complex Decision Systems, Chapter 4. Paris: Atlantis Press. 2010.

[25] T. Mackel, J. Rosen, and C. Pugh, "Data Mining of the E-Pelvis Simulator Database: A Quest for a Generalized Algorithm for Objectively Assessing Medical Skill", Studies in Health Technology and Informatics 119: 355-360, IOSPress, 2006.

[26] T. Mackel, J. Rosen, and C. Pugh, "Application of Hidden Markov Modeling to Objective Medical Skill Evaluation", Studies in Health Technology and Informatics 125: 316-318, IOSPress, 2007.

[27] P.B. McBeth et al., "Quantitative Methodology of Evaluating Surgeon Performance in Laparoscopic Surgery". Studies in Health Technology and Informatics, 85: 280-286, IOSPress, 2002.

[28] R. McCloy, R. Stone, "Science, medicine, and the future: Virtual reality in surgery", British Medical Journal 323(7318):912–915, 2001.

[29] R.M. Moraes and L.S. Machado, "Fuzzy Gaussian Mixture Models for On-line Training Evaluation in Virtual Reality Simulators", Annals Int. Conf. Fuzzy Information Processing (FIP'2003), 2: 733-740, 2003.

[30] R.M. Moraes and L.S. Machado, Hidden Markov Models for Learning Evaluation in Virtual Reality Simulators. International Journal of Computers & Applications 25(3):212-215, 2003.

[31] R.M. Moraes and L.S. Machado, "Using Fuzzy Hidden Markov Models for Online Training Evaluation and Classification in Virtual Reality Simulators", Int. Journal of General Systems, 33(2-3): 281-288, 2004.

[32] R.M. Moraes and L.S. Machado, "Maximum Likelihood for On-line Evaluation of Training Based on Virtual Reality", Proc. Global Congress on Eng. and Technology Education (GCETE'2005), pp. 299-302, 2005.

[33] R.M. Moraes and L.S. Machado, "Continuous Evaluation in Training Systems Based on Virtual Reality", Proc. Global Congress on Eng. and Technology Education (GCETE'2005), pp.1048-1051, 2005.

[34] R.M. Moraes and L.S. Machado, "Assessment Based on Naive Bayes for Training Based on Virtual Reality", Proc. Int. Conf. on Eng. and Computer Education (ICECE'2007), pp. 269-273, 2007.

[35] R.M. Moraes and L.S. Machado, "Fuzzy Bayes Rule for On-Line Training Assessment in Virtual Reality Simulators", Journal of Multiple-Valued Logic and Soft Computing 14(3-5):325-338, 2008.

[36] R.M. Moraes and L.S. Machado, "Online Training Assessment in Virtual Reality Simulators Based on Fuzzy Naive Bayes". Proc. Int. Conf. on Engineering and Technology Education (Intertech'2008), pp. 145-149, 2008.

[37] R.M. Moraes and L.S. Machado, "Development of Medical training Systems with Integration of Users Assessment". Book Chapter. In: Jae-Jin Kim (Ed.), Virtual Reality, Chapter 15. Croatia: Intech. 2011.

[38] D. Morris, C. Sewell, F. Barbagli, et al., "Visuohaptic simulation of bone surgery for training and evaluation", IEEE Computer Graphics and Applications, 26(6): 48-57. 2006.

[39] P.A. Pizzo, D.G. Poplack, Principles and Practice of Pediatric Oncology. Filadelfia, Lippincott-Raven Publishers, 1997.

[40] G. Riva, Applications of Virtual Environments in Medicine. Methods Inf. Med, 42: 524-534. 2003.

[41] G. Robles-de-la-Torre, "The Importance of the Sense of Touch in Virtual and Real Environments", IEEE Multimedia, 13(3):24-30, 2006.

[42] J. Rosen, C. Richards, B. Hannaford, and M. Sinanan, "Hidden Markov Models of Minimally Invasive Surgery", Studies in Health Tech. and Informatics 70:279-285, 2000.

[43] J. Rosen, M.C. Solazzo, B. Hannaford, and M. Sinanan, "Objective Laparoscopic Skills Assessments of Surgical Residents Using Hidden Markov Models Based on Haptic Information and Tool/Tissue Interactions", Studies in Health Tech. and Informatics 81: 417-423, 2001.

[44] J. Rosen, B. Hannaford, C. Richards, and M. Sinanan, "Markov Modeling of Minimally Invasive Surgery Based on Tool/Tissue interaction and Force/Torque Signatures for Evaluating Surgical Skills", IEEE Trans. Biomedical Engineering 48(5):579-591, 2001.

[45] J. Rosen, J. D. Brown, L. Chang, M. Sinanan, and B. Hannaford, "Generalized Approach for Modeling Minimally Invasive Surgery as a Stochastic Process Using a Discrete Markov Model", IEEE Trans. Biomedical Engineering, 53(3): 399-413, 2006.

[46] T. S. Sorensen, J. Mosegaard, "Virtual Open Heart Surgery – Training Complex Surgical Procedures in Cingenital Heart Disease", Proc. Siggrapgh Emerging Tech, paper 35, ACM, 2006.

[47] D.F.L. Souza, I.L.L. Cunha, L.C. Souza, R.M. Moraes, L.S. Machado, "Development of a VR Simulator for Medical Training Using Free Tools: A Case Study", Proc. Symposium on Virtual and Augmented Reality, pp. 100-105. Brazil. 2007.

[48] D.F.L. Souza, R.M. Moraes, L.S. Machado, "CybHaptics: A Class for Haptics Support in Virtual Reality Systems", Proc. of Symposium on Virtual and Augmented Reality (SVR2007), pp. 306-308. SBC. 2007.

[49] N. Taffinder et al., "Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator", Medicine Meets Virtual Reality. Amsterdam: IOS Press, pp. 124-130, 1998.

[50] T. Terano, K.Asai, and M. Sugeno. Fuzzy systems theory and its applications, Academic Press, 1992.

[51] W.S. Wilson et al., "MIST VR: a virtual reality trainer for surgery assesses performance", Annals of the Royal College of Surgeons of England 79:403-404, 1997.

[52] Y. Yang and G. Webb, "A Comparative Study of Discretization Methods for Naive-Bayes Classifiers", Proc. 2002 Pacific Rim Knowledge Acquisition Workshop (PKAW'02), pp. 159-173, 2002.

[53] L.A. Zadeh, "Fuzzy Sets", Information Control, 8: 338-353, 1965.

[54] L.A. Zadeh, "Probability Measures of Fuzzy Events", Journal of Mathematical Analysis and Applications, 10: 421-427, 1968.