A NEW CLASS OF ASSESSMENT METHODOLOGIES IN MEDICAL TRAINING BASED ON COMBINING CLASSIFIERS

Ronei Marcos de Moraes[1], Liliane dos Santos Machado [2]

Abstract *— Researches on training assessment for simulators based on Virtual Reality have less than 20 years old. By the use of simulators is possible to know users' performance during the training to analyze if they are prepared to perform the procedure in real situations. Basically, a Single User's Assessment System (SUAS) must continuously monitor all user interactions on VR environment and compare their performance with pre-defined expert's classes of performance to recognize user's level of training. In spite of the several methodologies proposed as kernel for SUAS, most of them are based on single classifiers. In this paper is discussed a new class of methodologies for SUAS based on Combining Classifiers.*

Index Terms — Medical Training, Users' Assessment, Virtual Reality, Combining Classifiers.

INTRODUCTION

Simulation systems based on Virtual Reality (VR) can provide realistic training for several areas and have been used since the 2nd World War, when the first flying simulators were developed [3]. In such immersive and interactive VR environments, which are replicas of real environments, users perform tasks that simulate real situations. It is interesting to know the quality of trainees' skills to perform those tasks in order to allow perform them in reality. Nowadays, in some areas, VR systems for training have been used to provide metrics to a proficiency criterion of learning, as in commercial aviation [1].

Researches on training assessment for simulators based on VR is a recent area for applications of Intelligent Decision Support Systems [5]. Assessment systems can work with qualitative or quantitative variables depending on the methodology used. Some can also use both types of variables [12,13] and the choice of the most appropriate method will depends on the problem addressed. That research area has less than 20 years old and from that, it is possible to know users' performance during the training to analyze if they are prepared to perform the procedure in real situations [13].

Probably the first works for Single User's Assessment System (SUAS) were proposed by Dinsmore et al. [4,7] using a quiz to assess user in a VR simulator to identify subcutaneous tumors. The quiz contained questions related to the diagnosis and hardness of tumor. Other research group [28] 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 [23]. These tasks are examples of systems in which a decision support system (DSS) can be coupled to the VR system [20].

Since 90's, several assessment methods were proposed [8,14,17,18,21], mainly for medical training. With the continuous increase of computers performance, SUAS could evolve too. Nowadays, a SUAS must continuously monitor all user interactions on VR environments 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 coupled or not 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. On the other hand, on-line SUAS are coupled to the training system and collect user data to provide a performance result at the end of the simulation [19]. An on-line SUAS works coupled to a VR simulator [14].

A SUAS should be capable to monitor user's interactions with the VR simulator by variables such as position (of user and of 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 shapes identified. All the information are sent to the SUAS which analyzes the data and emits, at the end of the training, an assessment report about the user's performance according pre-defined classes of performance.

Since the training simulators based on VR are capable of generating large volumes of data, it is interesting to design a knowledge modeling methodology based on numerical approach to extract from large data sets the expert's knowledge, allowing the design of a specific DSS. Thus, a SUAS is a complex computational structure, which has as a kernel a DSS. 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.

Among several methodologies proposed as kernel for SUAS, most of them uses a DSS, which generally are based on a single classifier. The main problem of that approach is

[1]Ronei Marcos de Moraes, Department of Statistics, Federal University of Paraíba - CCEN, Cidade Universitária s/n, João Pessoa – PB, Brazil, ronei@de.ufpb.br
[2]Liliane dos Santos Machado, Department of Informatics, Federal University of Paraíba - CCEN, Cidade Universitária s/n, João Pessoa – PB, Brazil, liliane@di.ufpb.br

the bias-variance dilemma [26]. A big sample of training data for parameters estimation of classifiers can reduce variance of its prediction. However, if that sample is not representative of data, it can induce a bias in the classifier's prediction.

According to Jain et al. [5], there are some advantages to use a combination classifiers in decision-making problem:

1. it can be used with a different number of classifiers based on different feature spaces simultaneously;
2. it can be used when more than one training data is available, which can contain different features;
3. it can use different classifiers on same feature space. Classifiers can have different performances in the same feature subspace;
4. some classifiers are initialized with random values and their performance can change due to it, even when is used the same training data. It can be interesting to combine several classifiers instead of taking the best and discarding the others.

Good results using multiple classifiers combined, improving system performance in several areas, can be found in the literature. Despite of their qualities, few papers could be found about methodologies for SUAS based on combining classifiers. However, there is no record that those methodologies have used that theoretical aspects explicitly. This paper discuss a new class of kernel for SUAS, which is based on combining classifiers.

Training Assessment Based on Virtual Reality

Virtual reality systems for training can provide significant benefits over other methods of training, mainly in critical medical procedures. Special devices can improve the realism of a VR application by the exploration of users senses, as 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 can provide 3D sound; and touch can be simulated by the use of haptic devices [3].

In some cases, procedures are performed without any kind of visualization and the only information noticed is the touch sense provided by robotic devices with force feedback. These haptic devices can measure forces and torque applied during the user interaction and the data can be used for an assessment [8]. Examples of haptic devices are showed in the Figure 1.

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 see them) by the use of this kind of device [3]. This is especially interesting in medical applications to simulate proceedings in which visual information is not available. Particularly, some haptic devices have shape and handhold similar to medical tools and can be used to simulate them. In this paper, the VR simulator used for tests was the bone marrow harvest simulator.

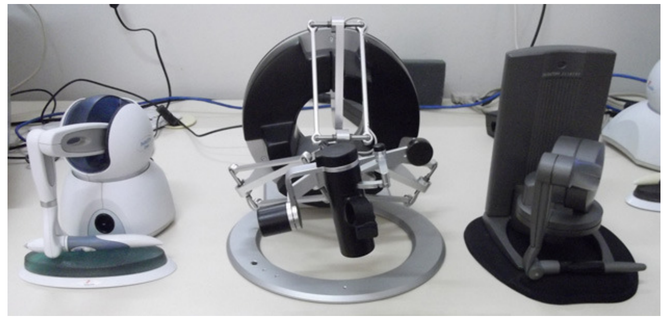

FIGURE. 1
Haptic devices with 6DOF.

Theoretical Aspects of Combining Classifiers

The research area of combining classifiers appears in the 90's of the twenty century. The main idea is the combination of classifiers in order to reduce deficiencies of a classifier by the good performance of others [6]. That combination is particularly effective when the classifiers are not similar to each other and none of them can solve the problem satisfactorily. Combining classifiers can be done by three architectures:

a) in sequential or linear way, when each classifier in the sequence improves results of previous classifier (Figure 2);

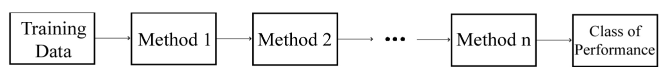

FIGURE. 2
Sequential Architecture for Combining Classifiers.

b) in parallel, when the results of each classifier are combined to provide the final result (Figure 3) and

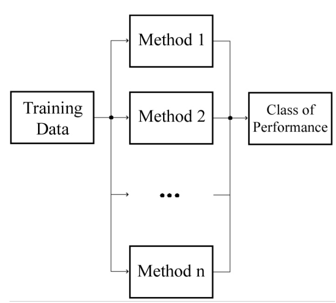

FIGURE. 3
Parallel Architecture for Combining Classifiers.

c) hierarchically, when both previous architectures can be combined to provide the final result (Figure 4).

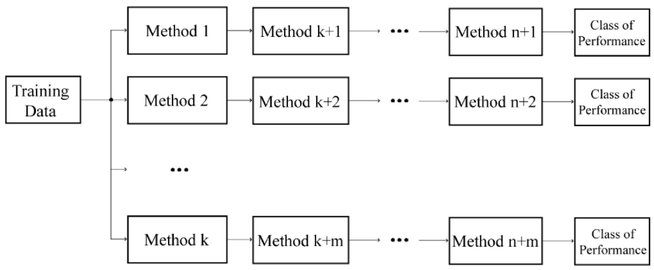

FIGURE. 4
HIERARCHICAL ARCHITECTURE FOR COMBINING CLASSIFIERS.

To provide the final decision, an architecture should be chosen. After that, it is necessary to choose the combination scheme of classifiers, which is called combiner [5]. Some characteristics of combiners must be known:

a) static or trainable: static combiners perform combination using a predefined rule and no training is required over the architecture, just over each classifier. The trainable combiners need it. In general, trainable combiners achieve better performance, but they need additional training data.

b) adaptive schemes: some schemes are adaptive and the combiner is able to evaluate the decisions of individual classifiers, taking into account the input data.

c) output information expected by combiner. Three levels are expected [27]:
   - c.1) measure or confidence: the classifier assigns a numerical value which expresses a belief or probability of a specific class given an input data;
   - c.2) rank value: the classifier assigns a rank to each class according to the input data;
   - c.3) abstract: the classifier assigns only a single class label, showing the best choice.

Several schemes can be found in the literature to combine classifiers, as voting, sum, mean, median, product, minimum, maximum, adaptive weighting, logistic regression, Dempster-Shapher theory and mixture of experts, among others [5].

In the following sections will be presented some approaches for SUAS found in the literature. Those approaches were used to perform assessment of psychomotor skills in medical training based on VR.

PREVIOUS WORKS USING COMBINING CLASSIFIERS

Formally, let be the classes of performance in space of decision $\Omega=\{1,...,M\}$ where M is the total number of classes of performance, which are degrees of psychomotor skills in medical training based on VR. Let be w_i, $i \in \Omega$ the class of performance for an user. Let be 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. $\boldsymbol{X}=\{X_1, X_2, ..., X_n\}$. The problem is identify which is the correct classe of performance w_i that $\boldsymbol{X}$ should be assigned.

Some approaches have been proposed using combining classifiers, although they did not mention it explicitly. Two of them proposed on-line SUAS for a bone marrow harvest simulator based on VR [9]. The third proposed an on-line SUAS for a gynecological examination simulator [11].

Gaussian Mixture Models + Relaxation Labeling

There are ways to improve the results of a classifier using another classifier after that. The Gaussian Mixture Models (GMM) is a good option for a SUAS based on virtual reality [16]. Formally, let a feature X_k from previous section, where $k=\{1,...,n\}$ for the class of performance w_i for an user. This feature can not assume a classical statistical distribution, but it can be modeled by a mixture of c Gaussian distributions. Thus, let $X_k=\{x_1, x_2,...,x_T\}$ be a set of T vectors obtained from a feature k, which is measure during the training for that w_i.

However, according to Tran et al. [25], the use of Relaxation Labeling (RL) after Gaussian Mixture Models can provide better results. The Relaxation Labeling (RL) was introduced by Rosenfeld et al. [22] and it is an interactive approach to update probabilities of a previous classification. This methodology is successfully employed in image classification. In this case, we will use RL after applied GMM classification. So, let be a set of objects $A = \{a_1,a_2,...,a_N\}$ and a set of labels $\Lambda=\{\lambda_1,\lambda_2,...,\lambda_N\}$. An initial probability is given to each object a_t having each label λ_k, which is denoted by $p_t(\lambda_k)$.

Moraes and Machado [15] proposed the composition GMM-RL as a kernel of a SUAS for virtual reality based environments (Figure 5) for a bone marrow harvest simulator. The procedure is performed without any visual feedback except the external view of the patient body, and the physician needs to feel the skin and other tissue layers trespassed by a needle to find bone marrow and then start its aspiration.

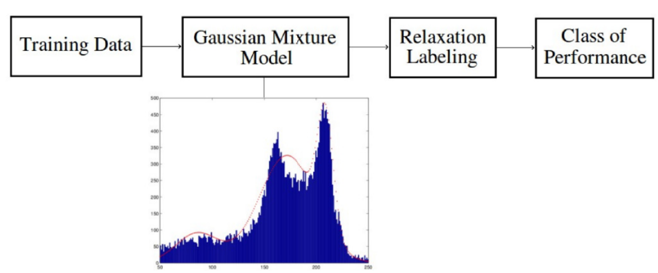

FIGURE. 5
DIAGRAM OF A SUAS FOR BONE MARROW HARVEST SIMULATOR.

Fuzzy Gaussian Mixture Models + Fuzzy Relaxation Labeling

In cases in which there is imprecision in variables measures or multiple ways for correct execution of a procedure, assessment can use a fuzzy approach. In those cases a fuzzy approach for GMM can be used. Tran and Wagner [24] proposed an approach named Fuzzy Gaussian Mixture Models (FGMM) that was followed by Moraes and Machado [14], as a kernel of a SUAS. FGMM uses a modification of Fuzzy C-Means algorithm [2] to estimate parameters: fuzzy mixture weights, fuzzy mean vectors and fuzzy covariance matrices.

Machado and Moraes [10] proposed the use of a two-stage SUAS where: FGMM is the first stage and Fuzzy Relaxation Labeling (FRL) is the second one. It is possible to consider previous results obtained from the FGMM method as initial degrees of membership to each assessment. The FRL method updates the membership degrees previously obtained and stops when the change in the degrees is less than a chosen threshold or equal to a chosen number of interactions.

Fuzzy rule-based expert systems

The simulator for gynecological exam (SITEG) [11] allows the training of gynecological exam and simulates different phases of pathologies. The simulator randomly presents normal, HPV/Herpes and inflamed cases to the user. The two stages of a real exam were divided to compose a visual and a touch exam. In the visual exam, the user must observe the vagina walls and cervix and notice their coloring. The evaluation system for SITEG is based on two fuzzy rule-based expert systems, each one according to the knowledge obtained from an expert, as presented in Figure 6. The first corresponds to the visual stage of the exam and the second corresponds to the touch stage.

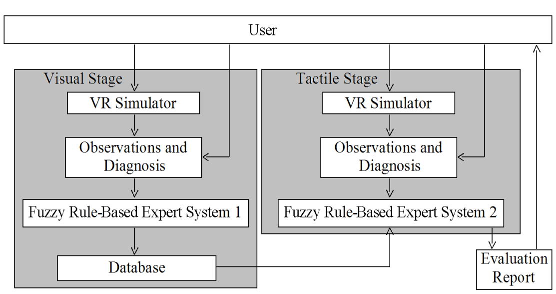

FIGURE. 6
DIAGRAM OF THE SUAS FOR GYNECOLOGICAL EXAMINATION SIMULATOR.

DISCUSSION

As mentioned above, despite of combining classifiers has been applied in several areas, few papers could be found in literature about methodologies for SUAS based its theoretical aspects explicitly. From papers found was possible to note that all proposed methodologies used sequential architectures with static and no adaptive combiners. The output information assigned by the first classifier was numerical value (probability or degree of membership). The second classifier refines that value to provide the final assessment.

It is possible to choose different architectures according to the features of the medical procedure simulated. It would allow using more sophisticated combination scheme of classifiers, as trainable and adaptive. As consequence, a new class of kernel for SUAS, which is based on combining classifiers (Figure 7) would help to design new an more accurate assessment methodologies. It is a new front of research in area of assessment systems for decision-making support.

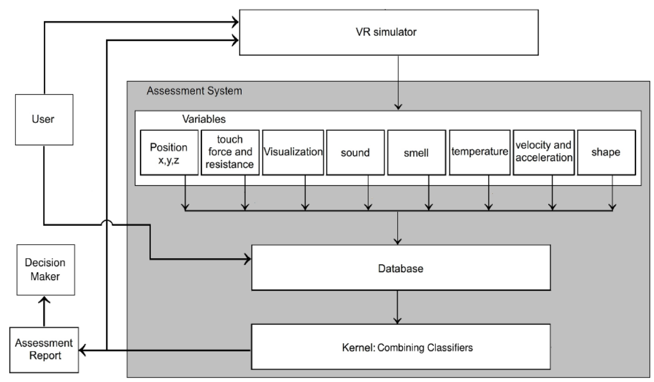

FIGURE. 7
DIAGRAM OF THE SUAS FOR GYNECOLOGICAL EXAMINATION SIMULATOR.

ACKNOWLEDGMENT

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] Aschwanden, C. et al., Centralized Data Recording for a Distributed Surgical Skills Trainer to Facilitate Automated Proficiency Evaluation. Studies in Health Technology and Informatics 125: 28-30, 2007.

[2] Bezdek, J.C. A review of probabilistic, fuzzy and neural models for pattern recognition. Journal of Intelligent and Fuzzy Systems, 1(1): 1-26, 1993.

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

[4] Dinsmore, M. VR simulation: training for palpation of subsurface tumors. Master's Thesis, Dep. Mechanical and Aerospace Eng., Rutgers University, October, 1996.

[5] Jain, A.K. et al. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1): 4-37, 2000.

[6] Kittler, J. et al. On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3): 226-239, 1998.

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

[8] Machado, L.S. et al. Fuzzy Rule-Based Evaluation for a Haptic and Stereo Simulator for Bone Marrow Harvest for Transplant. 5th PUG Workshop Proc., USA, 2000.

[9] 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.

[10] Machado, L.S. and Moraes, R.M. An Online Evaluation Of Training in Virtual Reality Simulators Using Fuzzy Gaussian Mixture Models and Fuzzy Relaxation Labeling. Proc. IASTED International Conference on Computers and Advanced Technology in Education (CATE'2003), pp. 192-196. Greece, 2003.

[11] Machado, L.S.; Moraes, R.M. VR-Based Simulation for the Learning of Gynaecological Examination. Lecture Notes in Computer Science , 4282: 97-104. Springer, 2006.

[12] Machado, L.S.; Moraes, R.M. Qualitative and Quantitative Assessment for a VR-Based Simulator. Studies in Health Technology and Informatics 142: 168-173, IOSPress, 2009.

[13] Machado, L.S., Moraes, R.M. Intelligent Decision Making in Training Based on VR. In:Da Ruan(Org.) Computational Intelligence in Complex Decision Systems. Cap 4: pp. 85-123. Atlantis Press, 2010.

[14] Moraes, R. M.; Machado, L. S. Fuzzy Gaussian Mixture Models for On-line Training Evaluation in Virtual Reality Simulators. Annals of the International Conference on Fuzzy Information Processing (FIP'2003). China, 2: 733-740, 2003.

[15] Moraes, R.M. and Machado, L.S. Gaussian Mixture Models and Relaxation Labeling for Online Evaluation of Training in Virtual Reality Simulators. Proc. Int. Conf. on Engineering and Computer Education (ICECE'2003), Brasil, 2003.

[16] Moraes, R.M. and Machado, L.S. Online training evaluation in virtual reality simulators using Gaussian mixture models. Studies in Health Technology and Informatics, 94: 42-44, 2003.

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

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

[19] Moraes, R.M., Machado, L.S. Assessment of collaborative medical training based on virtual reality using Possibilistic Networks. Safety, Health and Environment World Congress (SHEWC'2011), July, Santos [CD-ROM], 2011.

[20] Moraes, R.M.; Rocha, A.V.; Machado, L.S. Intelligent Assessment Based on Beta Regression for Realistic Training on Medical Simulators. Knowledge-based Systems. To appear.

[21] Rosen, J., Solazzo, M., Hannaford, B. and Sinanan, M. Objective Laparoscopic Skills Assessments of Surgical Residents Using Hidden Markov Models Based on Haptic Information and Tool/Tissue Interactions. Studies in Health Technology and Informatics 8:417-423, 2001.

[22] Rosenfeld, A. et al. Scene labeling by relaxat11ion operations. IEEE Transactions on Systems, Man and Cybernetics, 6(6): 420-433, 1976.

[23] Scott D.J. 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.

[24] Tran, D. and Wagner, Fuzzy approach to Gaussian mixture models and generalised Gaussian mixture models. Proc. Computation Intelligence Methods and Applications (CIMA'99), 154-158, USA, 1999.

[25] Tran, D.; Pham, T. and Wagner, M. Speaker recognition using Gaussian mixture models and relaxation labeling. Proc. 3rd World Multiconf. on Systemetics, Cybernetics and Informatics/ 5th Int. Conf. Information Systems Analysis and Synthesis (SCI/ISAS99), 6: 383-389, 1999.

[26] Tumer, K.; Ghosh, J. Alanysis of decicion boundaries in linearly combined neural classifiers. Pattern Recognition 29(2): 341-348, 1996.

[27] Xu, L et al. Methods for Combining Multiple Classifiers and their Applications in Handwritten Character Recognition. IEEE Transactions on Systems, Man and Cybernetics, 22(3): 418-435, 1992.

[28] 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.