

Performance Analysis of Assessment Using Maximum Likelihood for Virtual Reality Systems

Ícaro Lins Leitão da Cunha¹, Liliane dos Santos Machado², Ronei Marcos de Moraes³

Abstract — *Virtual Reality Systems (VRS) offer the possibility of creation of immersive environments for simulations of real life situations. The CyberMed is a free software system based in VRS which offers libraries for the development of simulation applications for medical teaching and training. To integrate an assessment functionality into the system, a class has been created based on the Maximum Likelihood classification method. This integration has been made in a way to guarantee a good performance of the class when using the system. The class presents two main features: to give a precise result based on comparisons between the user interaction and pre-established data and to provide that result in real-time. This work is based on the development of this class and on the analyzes of its performance when different scenarios are used, as the data size assessed and the system response time.*

Index Terms — *Assessment, performance analyzes, virtual reality systems.*

INTRODUCTION

In medical learning, surgery simulation is an essential tool for acquiring experience and ability. It can improve patient recovering time and reduce mortality indexes related to surgeries. Simulations based on virtual reality system (VRS) can offer an immersive environment that can be visually and physically (by means of specific interaction devices) realistic. In order to reach these characteristics, VRS depend on the quality of images in the environment, on the non-existence of delays during simulations and on the ways of interaction offered by the system [1]. The CyberMed is a VRS based on free software that offers a library of tools for developing simulation applications for medical teaching and training. Its main features are: three-dimensional visualization, use of realistic models, spatial interaction, visual sharing, and supervision/assessment of user interaction [2].

Assess user interaction during a simulation is important to know user's performance during the simulation and tell if, where and when he did something wrong during the simulation. From the user interaction the system will obtain interaction data and comparing it with pre-acquired data

from a specialist. This functionality will be developed as a class of the CyberMed system and it has two main objectives: to compute precise results and to return these results in real-time.

The data will be classified in classes, where each class represents a kind of performance of the user during the simulation. Depending of simulation, the data can be composed by one or more variable sets of data, where the number of variable sets represents the dimension of the class. Each variable set has a vector of data that can correspond to: position of the interaction device in one of the coordinate axis; interaction object's angle of approximation; forward force and rotation force executed on the device; etc.

Since the general goal of this work is to classify the user performance by using data from his interaction, the use of data classification methods had been necessary. In this case a method would be used as base of the conception of an assessment class. Additionally, this method should be precise in its classifications and return its results in real-time.

DATA CLASSIFICATION METHODS

Data classification methods are widely used in digital image processing (DIP). There are two kinds of classifying methods in DIP: visual classification, where images are interpreted by specialists; and the automatic classification methods, also known as numerical classification method. Numerical classification method can be supervised or non-supervised, where supervised methods are divided into two groups: classic methods and contextual methods. Since the goal of this work is to classify data from the CyberMed system, and not from images, and also due to the fact that the system needs pre-acquired data from a specialist, this work will focus on supervised classification methods.

Besides the fact that data need to be collected from the user, supervised methods also need a previous knowledge about the classes. It is the reason why these methods are more precise than unsupervised methods. In the case of the CyberMed, the assessment class uses coordinates, geometrical parameters and other information from a training stage to obtain values and other statistical characteristics of the classes. With these data it can execute

¹ Ícaro Lins Leitão da Cunha, scientific initiation student financed by CNPq, Department of Informatics, Universidade Federal da Paraíba – UFPB, Cidade Universitária s/n, João Pessoa, PB - Brazil, ivellius@yahoo.com.br.

² Liliane dos Santos Machado, Department of Informatics, Universidade Federal da Paraíba – UFPB, Cidade Universitária s/n, João Pessoa, PB – Brazil, liliane@di.ufpb.br.

³ Ronei Marcos de Moraes, Department of Statistics, Universidade Federal da Paraíba – UFPB Cidade Universitária s/n, João Pessoa, PB – Brazil, ronei@de.ufpb.br.

TRAINING AND ASSESSING

the classification of a simulation. In order to assure a good performance to a supervised classification it should follow two factors: the representation of estimations of numbers and statistical characteristics of the classes that belongs to the simulation; the degree of precision of the techniques that the classification is based on.

The Maximum Likelihood Method (MLM) is the technique chosen to be the basis of the assessment class. Prior studies of this method indicate it can offer a good performance and that it can generate results in real-time [3]. The MLM is based on the *Bayes* rule, in this case it will be used to classify a set of data.

Let $C = (C_1, C_2, \dots, C_n)$ be a set of classes, where n is the number of classes. For a one-dimensional data x , the probability that x belongs to the class C_i is $P(C_i|x)$, where $i = 1, 2, \dots, n$. If $P(C_i|x)$ is known for each class, then the class that x belongs to can be determined. This can be done by comparing all $P(C_i|x)$, but this value cannot be obtained directly, so the Bayes theorem should be used:

$$P(C_i|x) = P(x|C_i).P(C_i)/P(x)$$

where $P(C_i)$ is the probability of the class C_i to occur in a sample of data and $P(x)$ is the probability of x to occur in the sample.

Since comparisons done between $P(C_i|x)$ and $P(C_j|x)$, where $i \neq j$ (to determine to which class x belongs to) there is no need to calculate $P(x)$. So $P(x|C_i)$ and $P(C_i)$ are the only values needed to be obtained. From the Gaussian distribution formula for one-dimensional data:

$$P(x|C_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{(x - \mu_i)^2}{2\sigma_i^2}\right\}$$

where μ_i is the mean of the class C_i and σ_i its standard deviation. For multidimensional data [4]:

$$P(x|C_i) = \frac{1}{(2\pi)^{n/2} \sqrt{|V_i|}} \exp\left\{-\frac{1}{2}(x - \mu_i)^T V_i^{-1} (x - \mu_i)\right\}$$

where μ_i is the mean vector and V_i its covariance matrices.

To simplify the computation needed, the logarithm of $P(x|C_i).P(C_i)$ is used to obtain:

- For one-dimensional data: [5]
 $g(x) = -\log(\sigma_i) - \{(x - \mu_i)^2 / (2\sigma_i^2)\} + \log(P(C_i))$
- For multi-dimensional data:
 $g(x) = -\log|V_i| - (x - \mu_i)^T V_i^{-1} (x - \mu_i)$

When the value of $g(x)$ (which is called the evaluation function) is calculated for each class, the class to which x belongs to can be determined.

The assessment class was developed to integrate the CyberMed system. It is composed by two main routines: the training routine and the evaluation routine. In this case the assessment class stores two lists amongst its functionalities: the data class list and the interaction data list.

The data class list has information about the classes obtained during the training stage of the system. Some of the data stored in this list are: a $n \times n$ covariance matrix (where n is the class dimension); the determinant and inverse of this matrix; the mean vector (of size n) of the data of the class; and the class id. Data like the determinant and inverse are stored in a file. It prevents the class to execute calculation during simulation. This feature provides an important economy of time. The interaction data list stores the data that will be used either by the training stage (for obtaining the class data given by a specialist) or by the evaluation stage (in this case, obtained from user interaction that will have his performance evaluated). Among the data stored in this list there is the data vector of size $n \times m$ (where the size of m depends on the amount of interaction and the rate of storage of the data).

The file that stores the class data is divided in fields of the following format:

```
#ClassID ...
#MeanVector {
...
}
#CovarianceMatrix {
...
}
#CovarianceMatrixDeterminant {
...
}
#CovarianceMatrixInverse {
...
}
```

Each field in the file belongs to one class and these fields are composed by subfields that stores the data obtained during the training stage.

The training stage of the assessment class is executed by a specialist that will indicate the information about the data classes. This training is necessary for doing the assessment, because it will need the data to compose a series of discriminate functions that will classify the data during simulation. During this stage, each class will indicate a type of user performance during the simulation (bad, normal, good etc.). It is necessary to have at least one training execution for each class. During each training execution, the assessment class will create the list of data obtained from the specialist's interaction and from this the data class list is created and stored on a file that will be used during the assessment stage.

The assessment stage is executed during the system simulation. In this stage the system reads and stores the information on the data class file and obtains the user's

interaction data during the simulation. When the simulation is finished, the assessment class will classify and provide in real-time the result about user's performance.

INTEGRATION OF THE ASSESSMENT CLASS IN THE CYBERMED

The assessment class (CybAssess) is responsible for determining the user performance during simulation on the CyberMed system and it should give the results in real-time. In order to reach this, the integration of the CybAssess with the rest of the CyberMed classes must be done to guarantee that the communication between them will enable to store the interaction data without compromise the system performance.

The Figure 1 presents the integration of CyberMed classes used to collect data and to the assessment. The CybInterator class is responsible for obtaining values about the position and interaction variables collected from the interaction object in the simulation environment. These data are obtained either from the CybHaptics class (haptic device class) as from the CybMouseInteraction class (mouse device interaction class). The CybParameters class is responsible to provide other data that might be useful for the assessment, as geometric transformation parameters or ambient light intensity, as examples.

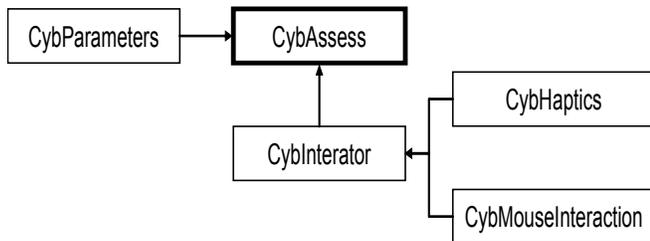


FIGURE 1

INTEGRATION OF THE CYBASSESS WITH THE OTHER CLASSES OF THE CYBERMED.

CASE STUDY

In order to test the assessment class performance, there was made a series of simulations with class of different dimensions. The main goal of these tests was to know accuracy and assessment capacity of CybAssess to perform assessment stage in real-time. Two sets of data samples where created for each test: a set for a training stage of the class and a set for an assessing stage. For each set 5 classes were defined with their respective variables set. The classes represent each category of users' performance and the variables set, in this case, the interaction variables to be analyzed. For each variable, a vector of size of 30000 samples was randomly generated from Gaussian probability distribution. The order of intersection between the variables set in each dimension is around 30% for these

probabilistically generated variables. This amount of intersection would be a good case to test the cybAssess accuracy. From these vectors a total of 10000 were used on the training stage, and the rest of the samples were used during the assessment stage. Table I presents a comparison of response time during the assessment stage for 5 classes where in each case presents a different variable set size (size equal 2, 3, 4 and 5). The amount of data used for the assessment stage is equivalent to a 20 seconds simulation with a haptics device frequency set at its maximum (at 1000 data/second). In this case, the simulation is being done at a worst case scenario for the assessment class. This amount of time is equivalent to the average amount taken in marrow collection procedure simulation [8].

TABLE I

AVERAGE TIME COMPARISON OF THE PERFORMANCE TESTS MADE ON THE CYBASSESS WITH 5 DATA CLASSES WITH A VARIABLE DIMENSION.

Dimension (variable set)	Average Time	Deviation (number of simulations = 40)
2	0.760 seconds	0.047
3	0.954 seconds	0.074
4	1.191 seconds	0.093
5	1.434 seconds	0.109

Figures 2 to 6 presents comparison charts of the assessment precision. In each case the number of variable sets in the classes is different. Each chart shows the variation of the precision (calculated using the comparison coefficient Kappa [6][7]) when the variable set size also varies. By the comparison of the 5 charts, can be noticed a growth in the precision of the assessment class when the data class dimension becomes higher.

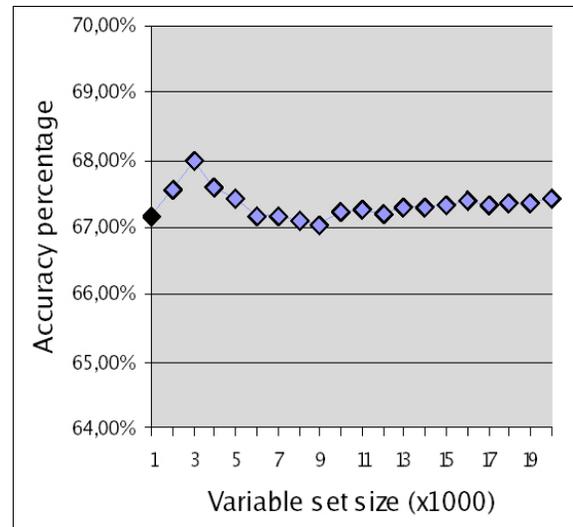


FIGURE 2

COMPARISON CHART OF ASSESSMENT ACCURACY WHEN VARIABLE SET SIZE VARIES FOR CLASSES WITH VARIABLE SET DIMENSION EQUAL TO ONE.

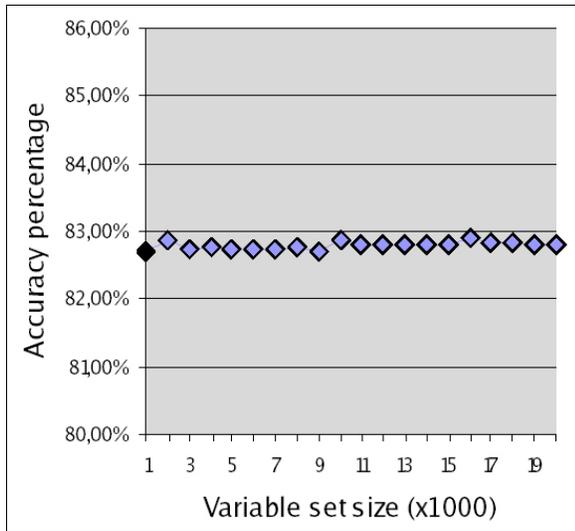


FIGURE 3

COMPARISON CHART OF ASSESSMENT ACCURACY WHEN VARIABLE SET SIZE VARIES FOR CLASSES WITH VARIABLE SET DIMENSION EQUAL TO TWO.

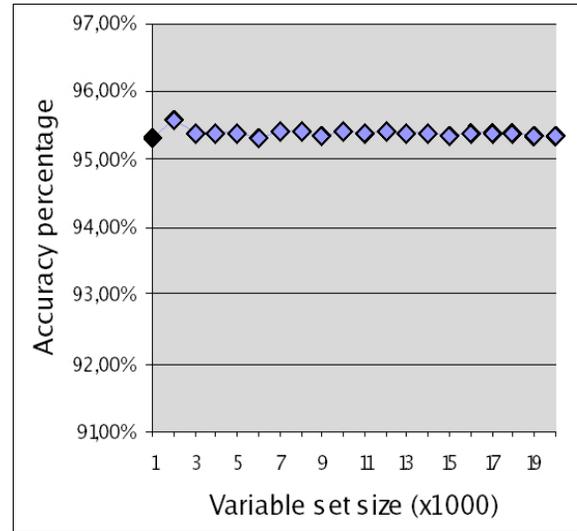


FIGURE 5

COMPARISON CHART OF ASSESSMENT ACCURACY WHEN VARIABLE SET SIZE VARIES FOR CLASSES WITH VARIABLE SET DIMENSION EQUAL TO FOUR.

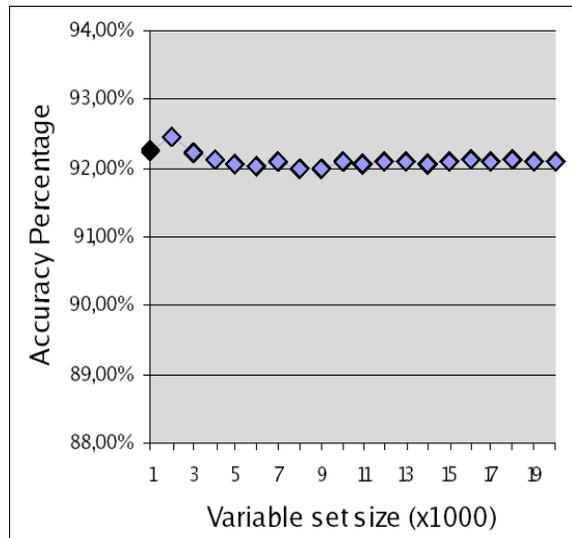


FIGURE 4

COMPARISON CHART OF ASSESSMENT ACCURACY WHEN VARIABLE SET SIZE VARIES FOR CLASSES WITH VARIABLE SET DIMENSION EQUAL TO THREE.

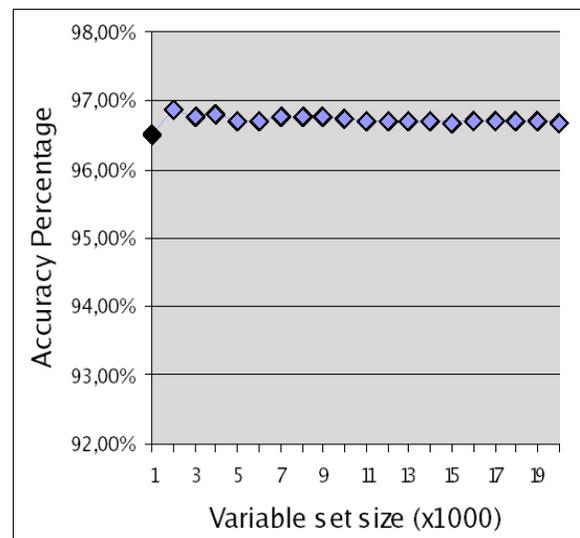


FIGURE 6

COMPARISON CHART OF ASSESSMENT ACCURACY WHEN VARIABLE SET SIZE VARIES FOR CLASSES WITH VARIABLE SET DIMENSION EQUAL TO FIVE.

When the assessment class is analyzed, a satisfactory performance can be noted in the average result return time and in the accuracy obtained in the test simulations. In this case, is important to observe that the class has to access the interaction data from the user and execute the discriminate function on them to calculate user performance throughout the whole execution of the simulation application. These test were made on a PC AMD64 3.2 Ghz with 1.0Gb of RAM running Fedora Core 4 32 bits.

ACKNOWLEDGMENTS

This work is partially supported by Brazilian Council for Scientific and Technological Development, CNPq (Process 303444/2006-1 and process CT-INFO-CNPq 506480/2004-6) and Brazilian Research and Projects Financing, FINEP (Grant 01-04-1054-000).

REFERENCES

- [1] Machado, L.S.; Cunha, I.L.L.; Campos, S.F.; Moraes, R.M., "CYBERMED: Realidade Virtual para Ensino Médico", *IFMBE Proceedings*. Vol. 5. 2004, pp. 573-576.
- [2] Basdogan, C.; De, S.; Kim, J.; Muniyandi, M.; Kim, H., et al., "Haptic Minimally Invasive Surgical Simulation and Training", *IEEE Computer Graphics and Applications*, Vol. 24, No. 2. March-April 2004.
- [3] Moraes, R. M.; Machado, L. S. "Maximum Likelihood for On-line Evaluation of Training Based on Virtual Reality". *Proceedings of Global Congress on Engineering and Technology Education (GCETE'2005)*. Santos, Brasil, 2005, p.299-302.
- [4] Johnson, R. A.; Wichern, D. W., *Applied Multivariate Statistical Analysis*. Prentice Hall, 4th ed., 1998.
- [5] Duda, R.; Hart, P.; Stork, D. "Pattern Classification". *Wiley-Interscience*. 2nd ed., 2000.
- [6] Brites, R.S.; Soares, V.P.; Ribeiro, C.A.A.S., "Comparação de desempenho entre três índices de exatidão aplicados a classificações de imagens orbitais. In: *Simpósio Brasileiro de Sensoriamento Remoto*, 8, Salvador, 1996. *Anais*. São Paulo: Image Multimídia, 1996.
- [7] Cohen, J., A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v.20, p.37-46, 1960.
- [8] Machado, L. S., Mello, A. N., Lopes, R. D., Odone Filho, V. and Zuffo, M. K. "A Virtual Reality Simulator for Bone Marrow Harvest for Pediatric Transplant". *Studies in Health Technology and Informatics*. Vol. 81. 2001, pp.293-297.