Graphical Representation as a Factor of 3D Software User Satisfaction: A Metric Based Approach

Apostolos Ampatzoglou[1], Alexander Chatzigeorgiou[2], Ioannis Stamelos[1]
[1]*Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[2]*Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece*

Abstract

During the last few years, an increase in the development and research activity on 3D applications, mainly motivated by the rigorous growth of the game industry, is observed. This paper deals with assessing user satisfaction, i.e. a critical aspect of 3D software quality, by measuring technical characteristics of virtual worlds. Such metrics can be easily calculated in games and virtual environments of different themes and genres. In addition to that, the metric suite would provide an objective mean of comparing 3D software. In this paper, metrics concerning the graphical representation of a virtual world are introduced and validated through a pilot experiment.

Introduction

Since 3D applications are of great interest for software industry and scientific community [6, 11, 21], the quality assessment of such applications can prove beneficial. Considering industrial purposes, it is obvious that the attractiveness of the product is proportional to its commercial success. In that sense, estimating software's attractiveness from the user's point of view can be used in sales estimation at an early development stage.

Even though there is a variety of scientific papers' concerning the quality assessment of 3D software, to the best of our knowledge, measuring user satisfaction through technical characteristics has not been addressed. This paper aims to estimate the product attractiveness to the user; it does not deal with evaluating the source code of 3D software with respect to complexity and maintainability [1 and 15]. According to [16], user satisfaction is a sub-characteristic of software usability. Additionally, considering that usability is one of the six software quality factors described in ISO/IEC 9126 [12], it becomes obvious that user satisfaction can be considered a factor of software quality.

In order to introduce the aforementioned technical characteristics, a systematic literature review has been performed according to guidelines presented in [4]. The results are thoroughly analyzed and the metrics that can be calculated through technical characteristics are extracted and validated through a pilot experiment. Pilot experiments (or pilot studies) are, according to Basili et.al. [2], important for the validity of the main experiment in the sense that the experiment scenario is confirmed and assistance in the experiment's organization is provided.

In the next section, an overview on the current state of the art in quality assessments of 3D software is presented. Next, the proposed metrics are introduced and analyzed. Later in the paper the pilot experiment that is used for metric validation is described and its results are presented. Finally, research limitations, conclusions and future work are discussed.

Previous Work

In this section of the paper, the current state of the art concerning the scientific research on assessing user's satisfaction from 3D software is being presented. From reviewing the literature it becomes obvious that most papers have either introduced metrics or heuristics in order to evaluate 3D worlds.

Firstly, in [10] it is suggested that computer game satisfaction factors are game genre related. The satisfaction factors that were under consideration (***Scenario***, ***Graphics***, ***Sound***, ***Game Speed***, ***Game Control***, ***Character*** and ***Community***) have been ranked according to their importance in several game genres (RPG, FPS, Sports and Boards). According to the paper, the most important factors have proven to be Character (20.0%), Graphics (17.8%) and Game Control (16.8%), while Community (10.1%) and Sound (10.8%) have appeared to be less important. A possible weakness of the paper lies in the fact that the authors do not describe the characteristics of a game with impressive graphics or solid characters.

Additionally, in [7] the authors have described fourty three (43) playability heuristics that were categorized with respect to ***Game Play***, ***Game Story***, ***Mechanics*** and ***Usability***. In [18], a model estimating players enjoyment based on flow has been introduced. The thirty six (36) heuristics used in the model are grouped in eight (8) categories: ***Concentration***, ***Challenge***, ***Player Skills***, ***Control***, ***Clear Goals***,

978-0-7695-3323-0/08 $25.00 © 2008 IEEE
DOI 10.1109/PCI.2008.12

Feedback, ***Immersion*** and ***Social Interaction***. In [13], the authors have extracted twenty-six (26) heuristics assessing the playability of mobile games that do not involve games mobility, but ***Game Play*** and ***Usability***. Finally, in [5] the authors have described the user preferences in FPS games concerning the ***Interface***, ***Map/Environment***, ***AI/Bot***, ***Multi-player Play*** and ***Single-player Play***.

In contrast to the extent that heuristics have been used in assessing 3D software quality, there are not many studies that have adopted the employment of metrics for the same reason. In [8 and 20], the authors attempted to depict users' experience of MOMRPG games through the network performance. The employed metrics are ***Jitter***, ***Packet Loss***, ***Ping***, ***Interactivity***, ***Consistency***, ***Network Fairness*** and ***Network Scalability***. Additionally, in [3] metrics describing immersion and presence in virtual environments are presented. The suggested metrics are ***Field of View***, ***Field of Regard***, ***Display Size***, ***Display Resolution***, ***Stereoscopy***, ***Head-Based Rendering***, ***Realism of Lighting***, ***Frame Rate*** and ***Refresh Rate***.

Evaluation Suite

In this section of the paper, considering that identifying metrics that influence all the variables described in [10] in one paper is not a trivial task, we have selected to firstly investigate the metrics that can be used in the prediction of the ***Graphics*** factor. The selection of this factor has been based on the fact that it is the second most crucial among the ones described in [10], and there is a strong belief that Graphical Representation is influencing two more variables, Character and Game Speed, which have not been investigating for correlation in the primary study.

By taking into account the findings of the literature review and our personal experience, the measurable technical characteristics (metrics) that are under consideration for involvement in estimating the value of the ***Graphics*** variable are: ***Average Number of Entities (NE)***, ***Average Size of Triangles (ST)***, ***Average Texture Size (TS)***, ***Average Texture Effects (TE), Number of Materials (NM), Average Number of Lights (NL), Average Environmental Effects (EE), Average Resolution Width (RW)*** and ***Average Resolution Height (RH)***.

Every 3D scene is a composition of *objects*. The objects that can take place in a scene are 3D geometries, textures, materials and lights. The average number of such objects in every scene is described by the *NE* metric.

A 3D geometry, commonly called *mesh*, is a set of vertices, grouped in triangles, which represent the general shape of an object. From the aforementioned description it becomes obvious that as the size of the triangles decrease, the smoothness of the object and therefore the precision of its representation increase. The *ST* metric is measured through the percentage of the average triangle size by the 3D mesh size.

Additionally, each vertex of the mesh is coloured in a specific way. The most common graphic programming APIs (openGL and direct-X) provide the designer with the ability to choose between colouring using textures and materials. *Textures* are images that aim to depict the details of an object. For example, when modelling a 3D car, matching a side image of the real car to the 3D model, makes it more realistic, in the sense that details such as scratches, colour noise etc are included in the scene. The *TS* metric value is normalized by dividing the physical average texture size with the value of 262144, which is a texture of size 512x512 pixels.

According to [17] there are several advanced texturing techniques that artists can employ in order to provide more realistic appearance to a scene. Such techniques are *bump mapping, light maps, opacity, specularity and illumination*. A technically "well-built" object is supposed to use at least one of those advanced texturing methods according to the object's appearance. The *TE* metric is calculated as the fraction of meshes using at least one texture effect by the total number of meshes.

The alternative to exclusive texture employment is the exclusive *material* use or the combination of the two colouring methods. In addition to colouring the model, materials are responsible for the calculation of lighting reflection. In that sense, every mesh in the scene is supposed to correspond to at least one material. The *NM* metric is calculated as the fraction of 3D geometries that are connected with materials by the total number of 3D geometries.

In this paragraph the possible environmental effects are investigated. According to [17], a 3D scene can use lighting, fog and shadows as special effects in order to enhance graphical quality. In respect to lighting, a possible metric is the average number of lights involved in every scene (*NL*). In addition to that, the *EE* metric is calculated as the sum of the fraction of scenes employing complete light calculations by the total number of scenes, plus the fraction of scenes employing fog by the total number of scenes, plus the fraction of scenes employing shadows by the total number of scenes.

Finally, it is expected that the size of the window displaying the 3D scene is analogous to user satisfaction. In that sense, the metrics *RW* and *RH* have been included in the metric suite for further investigation. The values of those metrics have also been normalized with respect to a display analysis of

640x480. The interested reader can access an overall picture of the mathematical formulas for the nine proposed metrics online in [19]

Pilot Experiment

This section of the paper aims at describing the experiment that was conducted, in order to validate the metrics introduced in the pervious section. The experiment has taken place according to guidelines described in [2] and is used as pilot for a formal experiment with higher number of subjects in a controlled environment. The results of the experiment can be used as indications for the metrics behaviour and as guidance in the definition and planning phase of the main experiment that will follow.

For this experiment thirty (30) testers have been asked to evaluate twenty four (24) screenshots of 3D scenes, with respect to graphics. The motivation for providing the testers with still images is to eliminate dependencies between their evaluation score and parameters other than graphics. This approach has also been used in [7] where the authors have validated the heuristics under study by screenshots that allowed users to navigate throughout the game but did not allow any game play. At this point it is necessary to clarify that since screenshots involve only one scene, the metrics' mathematical formulas have been simplified. The formulas used in the pilot experiment are available in [19].

The 3D scenes are created with a powerful 3D package, 3D studio max 7 (.max files), and the screenshots are produced by rendering a part of them in a jpeg format. The images have been selected or constructed, in order for the test set of the experiment to follow, as much as possible, the normal distribution with respect to every variable. In order for the manuscript to be more understandable and coherent, the graphs depicting the dataset deviation are excluded, but the interested reader can access them in [19]. The testers group is equally divided with respect to gender and is composed of users of various experience levels in gaming [19].

The testers have been given the dataset of images through an email and have been asked to subjectively evaluate their graphical quality (perceived graphical quality) in a one to twenty (0-20) scale. The users have also been asked to rank and review their answers before submitting, in order to avoid misjudgement in the image comparison. The testers have been allowed to equally evaluate different screenshots if they consider them of the same quality. After gathering the thirty (30) answers, the highest and the lowest scores have been excluded; the rest twenty eight (28) answers have been summed up and divided by twenty eight (28), in order to calculate the value of an Average Perceived Graphical Quality (*PGQ*) variable for each screenshot. The experiment variables, shown in Table 1, include nine ($1 \leq id \leq 9$) *independent* variables and one *dependent* (id=10).

id	Variable	Range
1	Number of Entities (NE)	≥0
2	Average Size of Triangles (ST)	0-1
3	Average Texture Size (TS)	0-16
4	Average Texture Effects (TE)	0-1
5	Number of Materials (NM)	0-15
6	Average Number of Lights (NL)	0-8
7	Average Environmental Effects (EE)	0-3
8	Average Resolution Width (RW)	0-5
9	Average Resolution Height (RH)	0-5
10	Average Perceived Graphical Quality (PGQ)	0-20

Table 1. Experiment Variables

The upper range of variables 3, 5, 6, 8 and 9, although theoretically unlimited, is set to the given value due to hardware limitation for rendering in real-time. For example, the values of *RW* and *RH* are limited to 5, after normalization, since current display devices rarely provide resolution greater than 3200x2400. Similarly, the *TS*, *NM* and *NL* metrics are limited to 16, 15 and 8 respectively, taking into account that textures greater than 8192x8192, meshes with more than 15 materials and scenes with more than 8 lights, are very unlikely to be properly rendered by common graphic devices in real-time. All metric scores greater than their upper limit are set to be equal to it.

Finally, most of the metric scores for the experiment's scenes have been calculated by manual observation of the (.max) files. In contrast to that, the calculation of the *ST* metric, that is highly unlikely to be manually calculated, has been made by a corresponding tool available in [19].

Experimental Results

As mentioned earlier, the pilot experiment aims at identifying indications on the significance of *metrics* in the calculation of the *Graphics* value. In order to achieve this task the experiment's data set, has been statistically analyzed. The employed techniques are ***Backward Linear Regression, Bivariate Correlation***, ***Two Step Clustering*** and ***Boxplots***.

The data set consists of 24 rows, each one representing one screenshot, and 20 columns. Each variable described in Table 1, is depicted in the data set by two columns, one with its numerical value and another with its value in a categorical scale. The recoding of numerical to categorical values has been made in order to permit greater flexibility in analyzing the data with techniques that cannot be performed with numerical values.

In order to examine the importance of each metric in the calculation of the **Graphics** variable, backward linear regression has been performed. Backward LR is an iterative process which excludes from the independent variable set, the least influential, with respect to the dependent variable, in each step. The most influential metric has proven to be *TE* (Texture Effects), closely followed by *NE*, *NL* and *TS*. On the other hand, *RH*, *RW* and *ST* have proven of minor importance. The created model is statistically important (sig.=0.001) and its fitness rate is quite satisfying (R^2=79.4% and adj. R^2=68.1%). The results of the regression are presented in [19].

In order to strengthen the aforementioned claim bivariate correlation tests have been performed. The difference from the previous approach is that at this time, each variable is individually examined for correlation with the dependent variable. This way, dependencies between independent variables are eliminated. The results are presented in Table 2 and are significant at the $p<0.05$ level.

Independent Variable	Pearson Correlation	sig
Number of Entities	0,571	0,002
Average Size of Triangles	-0,047	0,413
Average Texture Size	0,415	0,022
Average Texture Effects	0,599	0,001
Number of Materials	-0,310	0,070
Average Number of Lights	0,464	0,011
Average Environmental Effects	0,452	0,013
Average Resolution Width	-0,007	0,488
Average Resolution Height	-0,007	0,488

Table 2. Bivariate Correlation

From Table 2, it is implied that metrics *NE*, *TS*, *TE*, *NL* and *EE* are positively influencing the value of the ***Graphics*** variable. The negative values in the *Pearson Correlation* field suggest that as the corresponding metric's value increases, the value of the dependent variable decreases.

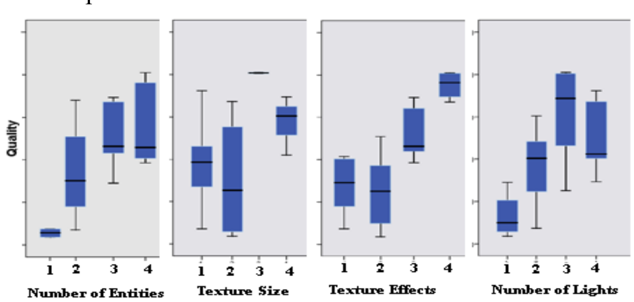

Figure 1. NE, TS, TE, NL Boxplots

In the above Figure (Figure 1), the relationship between *NE*, *TS*, *TE*, *NL* and the perceived graphical quality, is being graphically depicted through boxplots. As it is observed from the graphs, in general the quality of the scene improves as the value of each variable increases.

In order for the effect of *TE* metric to become clearer, in Figure 2 two screenshots that participated in the experiment are being presented. The two images are identical, with only one difference; the wall in the image on the right is using the bump mapping texturing technique in order to be more realistic (adding a small amount of noise and look more three-dimentional). The metric values are TE_{left}=0.4 and TE_{right}=0.6 respectively. Of thirty testers, six considered them of equal quality; twenty one considered the right image better and only three considered the left image better. In addition to that the quality values are PGQ_{left}=14.32 and PGQ_{right}=16.93, respectively. An important factor is that when many testers were asked to explain why they evaluated the right image with a higher score they haven't noticed the difference, but just felt the right one was better.

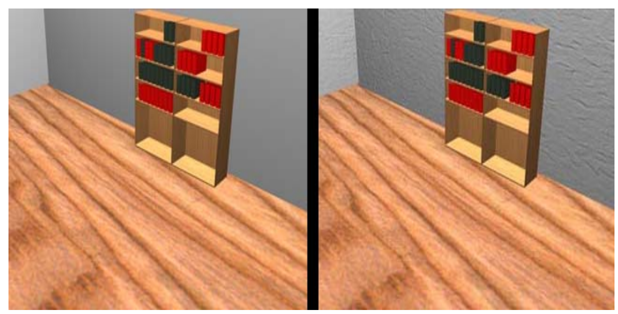

Figure 2. TE metrics – The bookcase example

Finally, in order to extract a profile for quality levels, two step clustering has been employed. The algorithm was asked to create five (5) clusters that fitted the five PGQ category scales ($PGQ_{(10-12)}$=1, $PGQ_{(12-14)}$=2, $PGQ_{(14-16)}$=3, $PGQ_{(16-18)}$=4, $PGQ_{(18-20)}$=5) and returned their centroids [19]. Such an approach suggests that according to the dataset, a 3D scene of the highest perceived quality is described by the following metric scores: NE=16, ST=0.09, TS=0.84, TE=0.83, NM=0.5, NL=2, EE=2 and RH=RW=1. According to the nature of clustering, scenes with greater metric values join the aforementioned cluster since their distance to its centroid is lower than of any other cluster.

Conclusions and Future Work

This paper aimed at introducing some technical characteristics of 3D software that could prove useful in estimating its perceived graphical quality, which is a major factor in user satisfaction. More specifically, nine satisfaction metrics have been introduced and described. In order to investigate the validity of those metrics a pilot experiment with 30 users and 24 3D scenes has taken place.

The results of the experiment showed that four of those metrics (***NE***, ***TS***, ***TE*** and ***NL***) are closely correlated to perceived graphical quality. On the other hand, the ***EE*** metric has proven to influence graphical representation, but not at the same degree. On the contrary, the rest of proposed metrics (***NM, ST, RH and RW***) have proven to be less important and therefore their definitions need reconsideration.

Concluding, the experience gained by this pilot study has proven extremely beneficial concerning the design of the formal experiment that will follow. More specifically, the ***RH*** and ***RW*** metrics are going to be merged in one metric (***Resolution***). Additionally, the ***ST*** metric is going to be reconsidered taking into account to more efficient mesh smoothing techniques. Furthermore, the need for higher number of testers and screenshots has been identified. Finally, each screenshot of the formal experiment is supposed to participate in the experiment with several variations with respect to the selected metrics under study.

At this stage of our work only metrics concerning the 3D application's graphics have been examined and validated. Furthermore, the study has not investigated the dependencies among the selected metrics. Consequently, a mathematical formula that combines the nine metrics and predicts the overall metric (PGQ) has not be calculated.

As future work, weights for each metric*,* that has proven to influence the graphics variable, should be estimated. In addition to that, metrics and weights for the other satisfaction factors should be introduced. By summarizing the aforementioned work, it might become possible to introduce an overall satisfaction metric.

Finally, it is necessary to clarify that in order to compare two 3D applications by taking into account their technical characteristics, it should be assumed that they are both developed by experts that have taken full advantage of the technologies they employ.

References

[1] A. Ampatzoglou, A. Chatzigeorgiou, "Evaluation of object-oriented design patterns in game development", *Information and Software Technology,* Vol. 49, Issue 5, May 2007, pp 445-454

[2] V. Basili, R.W. Selby, D. Hutchens. "Experimentation in Software Engineering", *IEEE Transactions on Software Engineering,* vol SE-12, No 7, July 1986.

[3] D.A. Bowman, R.P. McMahan, "Virtual Reality: How Much Immersion Is Enough?", *IEEE Computer*, Vol. 40, Issue 7, July 2007, pp. 36 - 43

[4] P. Brereton, B. Kitchenham, D. Budgen, M. Turner, M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain", *Journal of Systems and Software,* Elsevier, Volume 80, Issue 4, April 2007, pp 571-583

[5] D. Clarke, P.R. Duimering, "How Computer Gamers Experience the Game Situation: A Behavioral Study", *Computers in Entertainment,* Vol. 4, Issue 3, July 2006

[6] Communications of the ACM, *"Special Issue: Creating a science of Games"*, July 2007

[7] H. Desurvire, M Caplan, J.A. Toth, "Using Heuristics to Evaluate the Playability of Games", *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, Vienna, Austria, 2004, pp. 1509 – 1512

[8] S. Ferretti, C.E. Palazzi, M. Roccetti, G. Pau, M. Gerla, "FILA in Gameland, A Holistic Approach to a Problem of Many Dimensions", *Computers in Entertainment*, Vol. 4, Issue 4, Oct - Dec 2006

[9] T. Fullerton, "Play-Centric Game Education", *IEEE Computer*, Vol. 39, No. 6, June 2006, pp 36-42

[10] HyungBum Ham, YangSun Lee, "An Empirical Study for Quantitative Evaluation of Game Satisfaction", *Proceedings of the 2006 International Conference on Hybrid Information Technology (ICHIT '06),* pp. 724-729, November 9-11, 2006, Cheju Island, South Korea

[11] IEEE Computer, *"Special Issue on games education"*, June 2006

[12] H.W. Jung, S.G. Kim, C.S. Chung, "Measuring Software Product Quality: A Survey of ISO/IEC 9126", *IEEE Software,* Vol.21, Issue 5, September / October 2004, pp. 88-92

[13] H. Korhonen, E.M.I. Koivisto, "Playability Heuristics for Mobile Games", *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, Helsinki, Finland, 2006, pp. 9 - 16

[14] K. Mania, S. Ellis, M. Billinghurst, A. Steed, "Usability Evaluation Techniques for Virtual Reality Technologies", *Proceedings of the Virtual Reality Conference 2001 (VR'01)*, 13-17 March 2001, Yokohama, Japan

[15] D. Nguyen, S.B. Wong, "Design patterns in games", *Proceedings of the 33rd ACM SIGCSE Technical Symposium,* March 2002

[16] J. Nielsen, *"Usability Engineering"*, AP Profesional, *2003*

[17] M. Omernick, *"Creating the art of the game",* New Riders, 2004

[18] P. Sweetser, P. Wyeth, "GameFlow: A model for evaluating player enjoyment in games", *ACM Computers in Entertainment*, Vol. 3, No. 3, July 2005

[19] Usability of 3D software: A pilot Experiment - Webpage
http://sweng.csd.auth.gr/~apamp/pilot_usability.html

[20] A.F. Wattimena, R.E. Kooij, J.M. van Vugt, O.K. Ahmed, "Predicting the perceived quality of a First Person Shooter: the Quake IV G-model", *Proceedings of 5th ACM SIGCOMM workshop on Network and System Support of Games,* Singapore, October 2006

[21] M. Zyda, "Educating the Next Generation of Game Developers", IEEE Computer, Vol. 39, Issue 6, June 2006, pp. 30 – 34