Proceedings of the IASTED International Conference
HUMAN-COMPUTER INTERACTION
November 14-16, 2005, Phoenix, AZ, USA

TRIAL-AND-ERROR INTERFACE FOR TRANSFER FUNCTION SPECIFICATION IN DIRECT VOLUME RENDERING: A USER STUDY

Wen Qi
Department of Industrial Design
University of Technology Eindhoven
Eindhoven, The Netherlands
w.qi@tue.nl

Jean-Bernard Martens
Department of Industrial Design
University of Technology Eindhoven
Eindhoven, The Netherlands
J.B.O.S.Martens@tue.nl

ABSTRACT

Visualization via direct volume rendering is a very powerful technique for researchers to explore and interact with scientific data. However, the complexity of specifying a suitable transfer function (TF) for a specific data set causes many usability problems in the interaction process. The problems are such that they discourage the wider use of volume rendering. This paper presents an overview of important usability issues in the TF specification process, in order to derive an experiment in which some of these issues can be studied in a controlled way. More specifically, the experiment that we describe compares the user performance of some representative interfaces for TF specification. The experimental results have been analyzed from an engineering psychology point of view. The working memory has been identified as the core perception factor which has strong effects on the usability of TF specification. This is an important guideline for designers of such interfaces, and also suggests further questions that we intend to investigate in future experiments.

Categories and Subject Descriptors

I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism; H.5.2 [**Computer Graphics**]: User Interfaces

General Terms

Visualization

KEY WORDS

Volume rendering, transfer function specification, graphical user interface, usability testing

1. INTRODUCTION

Visualization via direct volume rendering is a powerful technique for exploring and manipulating large scientific data sets [6]. One problem that hinders effective use of it is the difficulty of understanding and specifying a TF for a specific data set, especially for non-expert users. The TF in a volume rendering system assigns optical properties, such as color and transparency, to the data values during the visualizing process. An appropriate TF can make a vast difference in quality and content of the rendered image. It is difficult to derive such a function automatically or manually as it is much dependent on the semantics of a specific data set. This paper discusses usability issues in TF specification, and analyzes the proposals that have been made in the literature to improve and optimize this interactive process. We summarize the current approaches in TF specification, and describe our visualization system prototype. Using this prototype, an experimental setup has been realized. We also present the results of our first usability test on the platform. We draw conclusions about technical and psychological aspects of the experiment, and describe our plans for future research activities in this area.

2. RELATED WORK

The TF is a critical component of the volume rendering process. It specifies the relation between scalar data (e.g. densities measured by CT or MRI imaging), and possibly also their first- or second-order derivative values, and optical characteristics, such as color and opacity [15]. Current graphics hardware-based algorithms provide the possibility to continually change the TF so that the results of volume rendering can be updated in real time. The analysis of the interaction process indicates that there are several steps involved in this TF specification (Figure 1). Ideally, the user could hope that the system provides sufficient information in the initial stage to finish the specification in a single step, as is indicated by the dashed arrow in Figure 1. However, the actual situation is that users usually need to go through multiple iterations of exploration and refinement before arriving at the final specification. During the initialization, a user is offered several inputs, such as derived data properties, like grey-value and/or gradient histograms, one or more initial TFs with correspondingly rendered images, etc. The user can explore the presented information and TF alternatives through a graphical or numerical user interface. He can asses the results of his operations based on the provided visual feedback. This visual feedback may not be restricted to the result of the last operation, but may also include feedback of preceding iterations and/or of the initialization step. The user refines his previous actions until he reaches his final goal, i.e., obtains a transfer function that results in a rendered image that adequately portrays the structure(s) of interest.

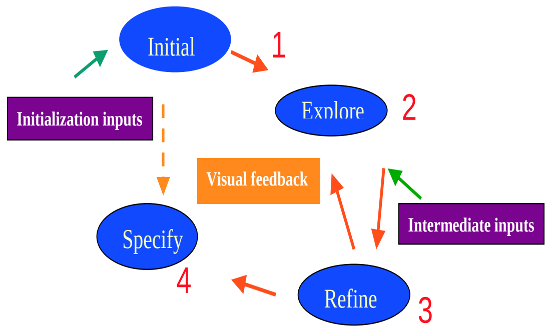

Figure 1: The iterative process of TF function specification.

The initial information that is presented by the system can consist of the following:

1. Data-dependent information such as histograms of grey values or (first- and second-order)derivatives of the input data, or a TF that is derived from the data through some sort of optimization algorithm;

2. Data-independent information that is based on prior knowledge or experience, such as standard or advised TFs (in medical applications, for instance, the TF may be determined by the kind of examination).

The intermediate feedback, in turn, can include the following:

1. Information from the initialization stage;

2. Visual feedback from the last operation of the user;

3. Feedbacks from one or more previous operations, that can assist in assessing the progress, without having to rely on memory.

The TF specification in volumetric visualization is a fairly unique and complex interaction compared to the elementary interactions, such as selection and positioning, that occur in most 3D graphics applications. It is only recently that this interactive process has become feasible in real-time, since it relies on the use of hardware graphic accelerators. Several alternative proposals have been made for how this interaction can be performed best [4, 5, 10, 11, 12, 13, 16, 17, 22, 24]. They range from completely manual to completely automatic, and differ in the amount and kind of feedback that is provided.

The most common method is the trial-and-error method. It involves manually editing the TF by modifying a graphical curve and/or by adjusting numerical parameters, and visually inspecting the resulting image (Figure 2. left) [21]. This method is primitive and problematic because it requires the users to go through all specification steps without intermediate feedback. Even with high-end facilities, this method can be very inefficient and time-consuming, because of the complexity of understanding the non-trivial relationship between a TF and the correspondingly rendered image. It also requires a reasonably accurate understanding of the visualization process by the user.

A method that tries to avoid the reliance on the user's visualization expertise is the Design Gallery [17] (Figure 2. right) approach. It involves creating and displaying a large number (hundreds) of rendered images that correspond to a range of predefined TFs. Design Gallery is an example of an image-centric method. Ma's image graph [16] and Kelly's spreadsheet [11] are related techniques. The image-centric methods do not focus on how to assist the user in finding a good TF by providing adequate feedback on relevant data-set properties, but instead focus on the design of the user interface. In the Design Gallery, all the user has to do is pick the rendered image icon that is most satisfactory, which implicitly selects the most suitable TF. The major challenge for this method is that possibly hundreds of volume renderings have to be created for the user to choose from. These random TFs need to be generated by the system such that they result in the widest spread of dissimilar output renderings. This implies that an automated way of judging dissimilarity is available, and the Design Gallery method hence has data-dependent characteristics through this dissimilarity measure. It is also not clear how reliably the user can judge the results of the alternative renderings based on the relatively small image icons. Because of the large number of image renderings that are required, Design Gallery also relies on real-time volume rendering functionality to be feasible.

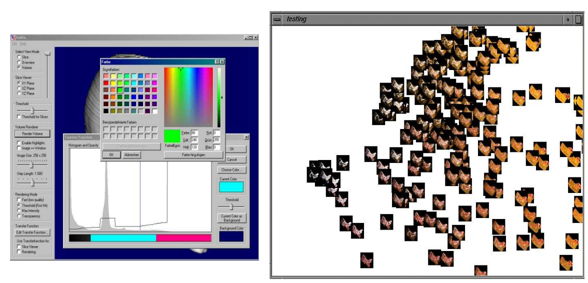

Figure 2: Two user interfaces for TF specification. Left: a trial-and-error interface; Right: The Design Gallery.

Kindlmann's semi-automatic method uses data-dependent properties to generate an optimized transfer function. It makes a reasonable assumption that the features of interest in a data set are the boundaries between different materials [12]. By making use of the relationship between the data values and their first and second derivatives along the gradient direction, Kindlmann's method can generate one solution for the TF from the multi-dimensional scatter plot of data values. It tries to remove the user from the interaction process and does not provide an intuitive interface. This method is very sensitive to noise and could not generate desired results for data with noise [21]. This automatic method is obviously data-dependent, but cannot be guaranteed to provide results that agree with user expectations. It may however be useful in the initialization stage. The automated method of Tzeng [24], on the other hand, uses a more intuitive interface and combines user input through a neutral network in order to select and adjust the TF. The user can for instance indicate areas in the rendered image that he finds interesting or not. It is a data-dependent method and achieved good results for one MRI data set. It is however not clear how their results extrapolate to other data sets. Their results can also not be reproduced, since the implementation details of their neural network are unknown.

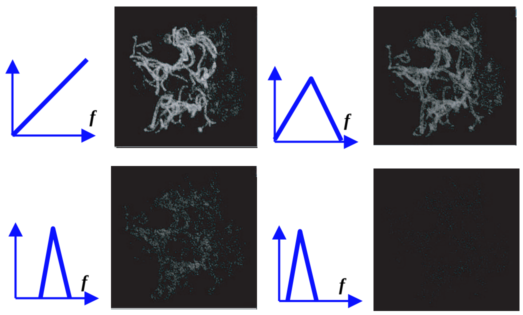

Figure 3: Some results of TFs versus rendered images for a data set containing an aneurism.

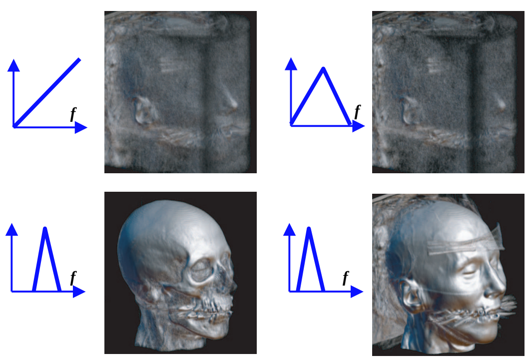

Figure 4: Some results of TFs versus rendered images for a CT scan data set of a head.

In summary, finding an appropriate TF can be described as a time-consuming and un-intuitive interaction task with all available methods. The main problems in TF specification that we have identified are the following:

1. *Too many degrees of freedom (DOF) in the interaction process.* As simple and direct as the principle of a TF is, it is extremely flexible as well, because of the immense variety of possible TFs. With the trial-and-error method, a user arbitrarily and repeatedly manipulates the coefficients of the mathematical representation of the TF, in order to adjust the visualization outcome. Common forms of such mathematical representations are piecewise linear functions or higher-order splines. Each control point in the graphical representation of a TF has two DOFs, because all control points are located in a 2D plane, so that even with a limited number of control points the number of DOF can be substantial. In the case of a TF specification that assigns opacity to grey values, and provided these grey values range from 0 to 255, the total number of DOFs is theoretically equal to 256.

   A user is usually guided in his interactions with the TF by how closely the rendered image matches his interaction goal, which is most often to adequately reveal specific structures in the data set. Slight changes in TF can however change the result dramatically. Moreover, the specification of the TF is data-set dependent. For different data sets, the same TF will not achieve similar results, as is illustrated by comparing Figures 3 and Figures 4. A TF that is well suited for one data set may be completely useless for another one. Therefore, even for a visualization expert it is a hard task. Most of the targeted users of visualization tools are moreover domain experts, that cannot be expected to have a deep understanding of the relationship between the TF and the rendered image.

2. *Inappropriate design of the user interface and inadequate information for TF specification.* The amount of control that a user has over the TF is determined by the interface. Besides the fundamental problem of the large number of DOFs, a poor design or arrangement of the user interface can make the TF specification more difficult and less efficient, especially when useful information is not available. This may be a contributing factor to why many of the available interfaces frustrate the user and fail. As we have seen, there have been several suggestions for solving this problem by creating more intuitive interfaces (for example, the Image Graph [16]). Most of them suggestions have however never been evaluated in a formal user study. Therefore, it is not clear which information is really useful to the end user, and which one is not. In the absence of such knowledge, we can expect that useless information is sometimes provided by interfaces, while valuable information may be missing.

3. EMPIRICAL WORK

The TF specification has been listed among the top ten problems in volume rendering [21], and we propose to use empirical research to get a better grip on this problem. The analysis of the interaction process, summarized in Figure 1, has indicated that a proper method for TF specification should support useful feedback. Improved solutions for TF specification are those that minimize the effort for the user from exploration to refinement. The trial-and-error method provides only minimal visual feedback, and we therefore consider it as our baseline system. We have devised an experiment by which we explore the usefulness of additional information in this trial-and-error method. More specifically, we aim to assess the effects of the following additional feedbacks: 1) data-dependent information, such as the histogram; 2) data-independent information, such as suggested or standard TFs. We also wanted to investigate the effect of a graphical user interface with a limited number of DOFs. We a priori formulated the following three hypotheses:

1. *Data-dependent (more specifically, histogram) information assists the users in TF specification.* Most available interfaces offer such information, so that it seems to be generally accepted that it can help the user in his search for a proper TF. The most frequently provided information is the grey-value histogram. It graphically depicts the frequency distribution of grey values in the data. In a histogram, the horizontal axis represents the range of grey values from 0 (shadows) on the left to 255 (highlights) on the right. In a stan-

dard histogram, the vertical axis represents the number (or percentage) of pixels that have each one of the 256 grey values. The higher the line coming up from the horizontal axis, the more pixels there are with that grey value in the data. In a cumulative histogram, on the other hand, the vertical axis represents the number (or percentage) of pixels that have a value smaller than or equal to each of the 256 grey values. The cumulative histogram integrates the standard histogram, and therefore has a more regular shape. We will use the cumulative rather than the standard histogram as the data-dependent feedback in our experiment.

2. *Data-independent (more specifically, suggested TF) information assists the users in TF specification.* It is supposed that data-independent information comprises suggestions for the user to narrow down his search. These suggestions can for instance be in the form of geometric shapes of TFs, such as triangular, rectangular, hat-shaped, level-up, up-level, and step-like functions. These suggestions are based on rendering experience. Although the TF is data dependent, there are simple TFs, such as piecewise-linear TFs, that often produce reasonable results. Higher-order spline representations of the TF are more difficult to control and seem to have only limited added value in most cases. Moreover, within the class of piecewise-linear TFs, not all shapes are equally likely to produce meaningful results, and the shapes that are expected to be most useful a priori can be suggested. Decreasing functions will for instance be absent from these suggestions, because they often do not create useless results.

3. *A graphical user interface with limited DOFs in the TF control assists the users in TF specification.* A user manipulates the transfer function via a graphical user interface. Because the main difficulty for TF specification is too many DOF, it is reasonable to think that users might have less difficulty if they are presented with an interface with limited DOFs. We therefore will also test the case where the piecewise-linear TFs that we describe above are not only provided as suggestions, but are actually the only shapes available to the user. In order to properly evaluate this case, we will obviously not only have to look at the time that people take to realize a TF setting, but also at the quality of the result that they produce.

3.1 Apparatus

In order to enable us to experimentally investigate the usability aspects of TF specification, we needed to create a volume visualization prototype. Our experimental hardware setup consists of a DELL graphics workstation (Pentium IV, 2.4 GHz, 512 MB RAM, FireGL 4 graphic card); a 17' CRT monitor; a 14' CRT monitor; a keyboard and a mouse; and loudspeakers (stereo). The key software component of the system is a volume-rendering engine that visualizes volumetric data with the help of hardware-accelerated 3D texture mapping [23]. This implies that the TF specification is implemented by means of a texture look-up table.

3.2 Interfaces

In the experiment, the visual feedback at any time consists of a single TF, with its available controls and feedback, and

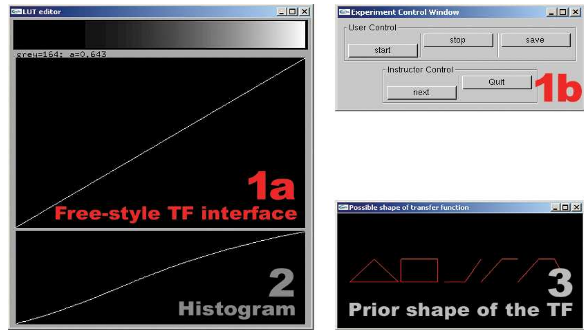

Figure 5: The user interfaces for experimental conditions 1 (part 1), 2 (part 1+2), 3 (part 1+3) and 4 (part 1+2+3).

the correspondingly rendered image of the scientific data set. The participants interact with the TF via a graphical user interface. The experiment involves five interface conditions. The five kinds of user interfaces represent different strategies corresponding to our hypotheses. The baseline interface with free-style control, referred to as condition 1, consists of parts 1a and 1b in Figure 5. With this free-style interface, a user has full control over the TF. The panel 1b controls the course of the experiment, i.e., starting and stopping a single TF control trial, saving the rendered image, or loading a new data set. The part 1a allows the user to manipulate the TF by creating and moving control points of a piecewise linear function along the horizontal and vertical direction within a 2D interaction area. There is no movement limitation for the control points except that the grey values for the first and last point have to remain at 0 and 255, respectively. The user can create TFs with as many control points as he wants. Experimental condition 2 includes data-dependent information, and consists of parts 1 and 2 in Figure 5. A cumulative histogram and free-style TF interface are presented at the same time. Experimental condition 3 includes data-independent information, and consists of parts 1 and 3 in Figure 5. In condition 4, both data-dependent and data-independent information are offered, so that all parts in Figure 5 are presented. The interface for our final condition 5 is shown in Figure 6. It is a user interface that allows for a number of piecewise-linear TFs, and that does not provide (data-dependent) histogram information. Each kind of TF is represented by a graphical icon. The available TF curves have only few control points and limited DOFs. This implies that the shape of a curve cannot be altered.

3.3 Procedure

There were 13 participants in the experiment, five female and eight male persons between 19 and 50 years old. All of them have university education in engineering or science. Each subject participated in all conditions. The order in which conditions were presented to the participants was randomized. The participants were given a consent form to read and sign.

Upon entrance, participants were given an experiment instruction sheet that described the system and tasks to be performed. These written instructions remained available during the entire experiment. The participants were introduced to all five user interfaces and could interact with them,

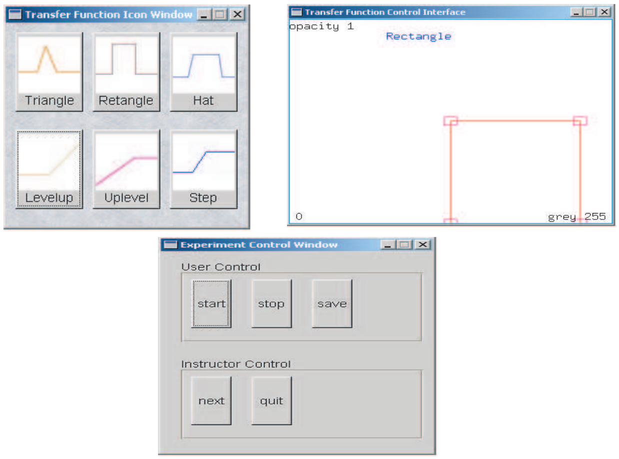

Figure 6: The user interface for experimental condition 5.

using a data set that was not part of the actual experiment. Afterwards, each participant performed four tasks in each of the five interface conditions. Each task involved a different data set, and required the participants to visualize a prescribed structure within the data set as good as possible. The order in which the four tasks were executed with an interface was also randomized across subjects. The dependent variables recorded during the experiment were the following:

- the total number of the mouse clicks during a task;
- the number of the mouse clicks for each icon in condition 5;
- the time needed to finish an task;
- the rendered image produced at the end of a task;

The participants were given several questions to answer after the experiment in order to collect their subjective impression of the interfaces and the produced images. The design of the questionaire was based on available usability questionaires [9], [14]. More specifically, the questionaires contained the following parts:

- Personal data, such as age, education, and former experience or knowledge about visualization systems and the problem of TF control;
- Their agreement or disagreement with general usability statements about the system, such as: "It was easy to use the system";
- Questions addressing the usability of the individual interfaces, more specifically, subjects were asked to rate effectiveness, efficiency, satisfaction and overall quality (the detailed questions will be specified later);
- In order to assess the task performance, which was defined as the degree to which the produced images matched the participants' goal, they were asked to both rate the output images individually and to express their preference for all pairwise combinations of output images [1, 8];
- Follow-up questions with an open answering format, in order to collect additional (a priori unexpected) comments from subjects.

All the subjective ratings of the subjects were performed on a 7-point scale.

4. RESULTS

In the following sections, we discuss the quantitative results of the experiment, as well as the results from the subjective evaluation by means of the questionaires.

4.1 Quantitative data

The Figure 7 shows the mean time for all five conditions and all four tasks. There are no significant effects in the "lobster", "head", and "foot" data sets. For the "engine" data set, the order of the conditions with respect to performance time is one to five from faster to slower. Further "Analysis of Variance"(ANOVA) with repeated measures was carried out on the mean time with a significance level of α =0.05. The results are the following:

Engine There is a significant effect of interface condition, F(4,52)=9.141,$p<.05$. The post hoc test shows that between condition 1 and 5, condition 2 and 5, and condition 3 and 5, there were significant differences. Condition 4 and 5 do not demonstrate a significant difference.

Foot There is no significant difference,F(4,52)=.472,$p>.05$.

Head There is no significant differences,F(4,52)=1.001,$p>.05$.

Lobster There is no significant differences,F(4,52)=.268,$p>.05$.

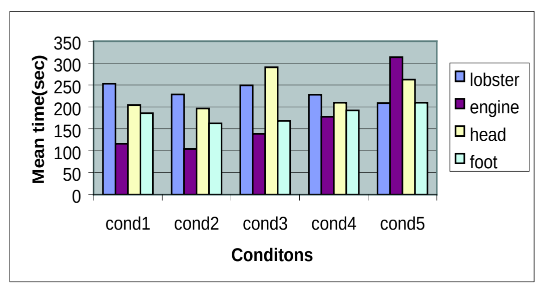

Figure 7: The mean time for five conditions with four tasks.

The Figure 8 shows the average number of mouse clicks for all five conditions and all four data sets. There were no significant effects in either of the four data sets. However, with the "lobster", "head" and "foot" data sets, there is a tendency for the number of mouse clicks to be lower in condition 5 than in the other four conditions. ANOVA with repeated measures on the average mouse clicks for four tasks shows the statistical details:

Engine There is no significant effect,F(4,52)=2.411,$p=0.061>.05$.

Foot There is no significant effect,F(4,52)=.774,$p=0.547>.05$.

Head There is no significant effect,F(4,52)=1.841,$p=0.135>.05$.

Lobster There is no significant effect,F(4,52)=2.202,$p=0.082>.05$.

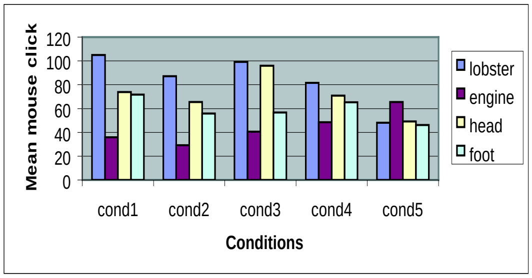

Figure 8: The mean mouse clicks for five conditions with four tasks.

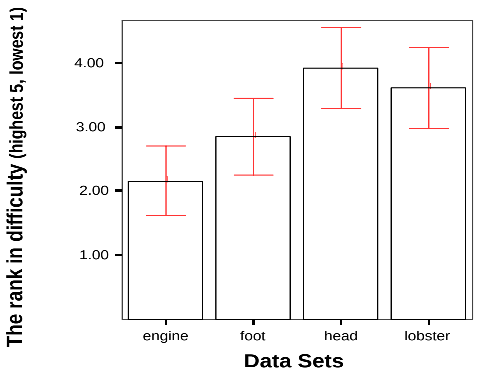

Figure 9: The average rank of difficulty for the four tasks (i.e., data sets). Bars show Means; Error Bars show 95.0% Confidence Interval for Mean.

4.2 Subjective Evaluation

This subjective evaluation is based on the answers of the questionaires. The results are mainly summarized along four important characteristics, i.e., effectiveness, efficiency, satisfaction and overall preference. The four tasks are also evaluated in difficulty in order to find out whether or not there is a correlation between the difficulty of the task and the user performance. These subjective evaluations are also compared against the quantitative results in order to test our hypotheses.

The difficulty of the tasks The tasks with the four data sets presented different difficulties to the subjects, as is shown in the figure 9. The task with the "engine" data set has been recognized as the easiest one. The most difficult task is the one with the "head" data set. The feedbacks from the subjects indicated some of the reasons. The task with the "engine" data set could be finished satisfactorily with a simple TF. Most of the subjects found that the task with the "head" data set required too much details to be rendered simultaneously. The differences in difficulty of the four tasks are assumed to reflect real situations and should therefore be taking into account in the experiment.

The task performance judging by the image quality In order to evaluate the effectiveness of the different interfaces, the task performance in terms of image quality is taken into account and measured. The image quality is scaled by how close the produced image compares with the original goal. Figure 10 illustrates that

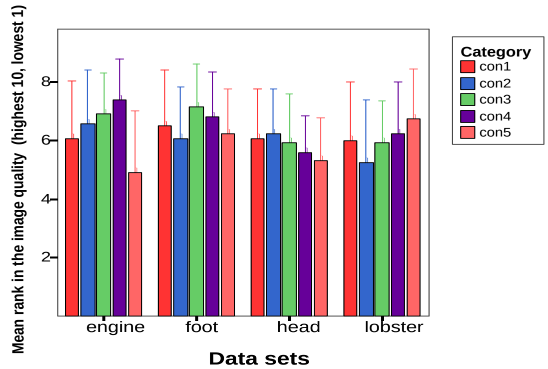

Figure 10: The average rank of task performance for the four tasks (i.e., data sets). Bars show Means; Error Bars show 95.0% Confidence Interval for Mean.

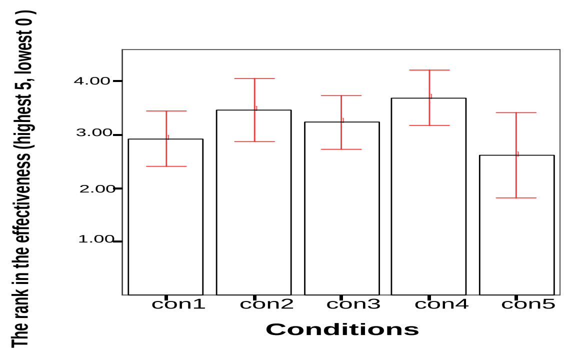

Figure 11: The subjective evaluation of effectiveness of five TF interfaces.
Bars show Means; Error Bars show 95.0% Confidence Interval for Mean.

with the "engine" and "head" data set, the task performance in condition 5 is the worst. However, the best performance has been achieved in condition 5 with the "lobster" data set.

Effectiveness In terms of effectiveness, we asked the subjects to evaluate "which condition or interface provided more control over the TF". Condition 4 get the highest rank, with Condition 2 as a close second. Condition 5 is considered least effective. Figure 11 shows the details.

Efficiency Efficiency was defined as "how fast the user thinks he has finished the task" using the different interfaces. Figure 12 upper right illustrates that condition 4 again scores best, while condition 5 is worst.

Satisfaction Look and feel is a very important factor in the design of an interface. We asked our users to give an evaluation on "the arrangement of the five interfaces". The results are illustrated in figure 13. Surprisingly, condition 4 still gets the highest appreciation. Condition 5 scores higher in this attribute than in the previous attributes.

Overall preference Considering all the three elements above, subjects gave their evaluation of the overall preference

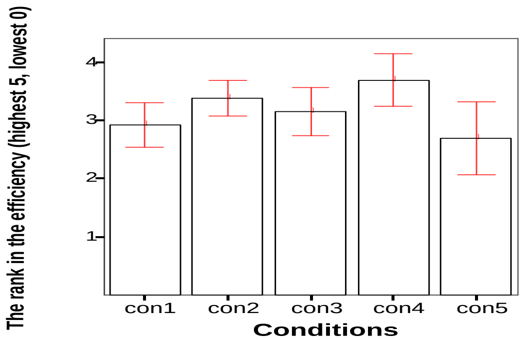

Figure 12: The subjective evaluation of efficiency of five TF interfaces.
Bars show Means; Error Bars show 95.0% Confidence Interval for Mean.

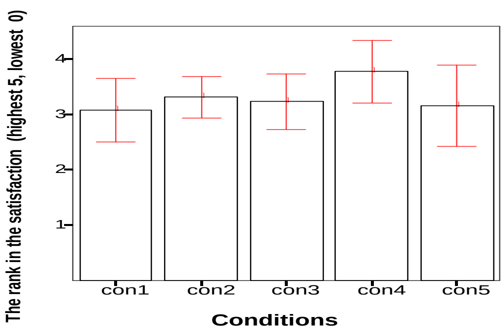

Figure 13: The subjective evaluation of satisfaction of five TF interfaces.
Bars show Means; Error Bars show 95.0% Confidence Interval for Mean.

of the five interfaces. The results are shown in Figure 14. The condition 4 is ranked highest. Condition 1,2,3 are very close to each other, while Condition 5 is least appreciated.

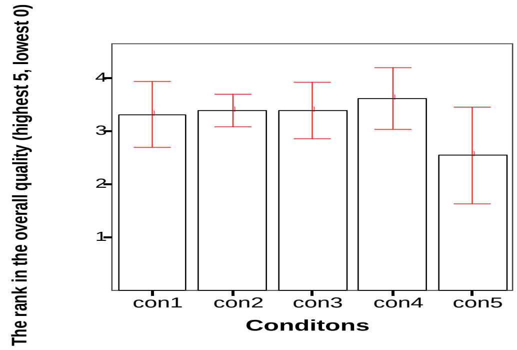

Figure 14: The subjective evaluation of overall preference of five TF interfaces.
Bars show Means; Error Bars show 95.0% Confidence Interval for Mean.

5. DISCUSSIONS

Although the experimental results do not provide strong evidence to support our three hypotheses, some valuable information could be found through them. We discuss the results and explain the reasons individually.

5.1 Without histogram vs. With histogram

Both quantitative and qualitative analysis of the results do not support the hypothesis that the histogram assists the users in the TF specification. Although most of the subjects could understand the principle and the purpose of the histogram, it is hard to apply it during the specification process. The strongly required mathematical background hinders the subjects to derive useful information from the histogram. Only two subjects who had experience with Photoshop and the histogram feature within it, gave positive feedback on the use of histogram. These subjects also used the information about grey value distribution that they derived from the histogram in the construction of their TFs. Therefore it seems that, for most subjects, the histogram does not assist in the TF specification, although the histogram is provided as a default setting in most of traditional volume rendering application.

5.2 Without additional TF information vs. With additional TF information

Contrary to our hypothesis, we did not find any statistically significant differences between these two conditions. The additional information does hence not help the users to speed up their exploration and refinement process. However, the users with less background knowledge on volume rendering still appreciated this information, and they preferred it to be present, since it did help to get them started. Since the information does not present an active mechanism to shorten the search time, it is not unexpected that it did not improve the user performance in terms of execution time.

5.3 Free-style vs. Limited DOF

Surprisingly, the graphical user interface with limited DOF in condition 5 did not help subjects in terms of time and mouse clicks. For the "engine" data set, which represented the easiest task, the mean trial time even increased. The mean task performance in terms of image quality is the best for the "lobster" data set, which is not the most difficult task. This implies that interface 5 might not be suitable for a very simple task, neither the most difficult task, for example the "head" data set. The comments from the subjects suggest that the limitations in DOF are mostly useful for exploring the data and accumulating experience. They are an obstacle in a task that requires the user to perform more subtle adjustments, such as in the case of the "head" data set. But it might be suitable for a task that has intermediate difficulty, which is the case for the "lobster" data set.

5.4 Working memory for TF specification

Further analysis through the knowledge of engineering psychology and perception could explain why such a specification task is difficult. Human beings have two different storage systems with different durations with human being: working memory and long term memory [25]. Working memory is the temporary, attention-demanding store that we use to retain new information until we use it [2]. Human beings use working memory as a kind of "workbench" of consciousness where we examine, evaluate, transform, and compare different mental representations. Human beings might use working memory, for example, to carry our mental arithmetic or predict what will happen if we schedule jobs one way instead of another. Finally, working memory is used to hold new information until we give it a more permanent status in memory-that is, encode it into long term memory.

However, several experiments have demonstrated the transient character of working memory [7]. Estimates generally suggest that without continuous rehearsal, little information is retained beyond 10 to 15 seconds. This transient character of working memory presents a serious problem for those work domains/tasks when information can not be rehearsed [20].

Working memory is also limited in its capacity (the amount of information it can hold) [3]. And this limit interacts with time. Experiments show that faster decay is observed when more items are held in working memory, mainly because rehearsal itself is not instantaneous [18]. The limiting case occurs when a number of items can not successfully recalled even immediately after their presentation and with full attention allocated to their rehearsal. The limiting number is sometimes referred to as the memory span. In a classical paper George Miller identifies the limits of memory span as the magical number seven plus or minus two [19]. Thus, somewhere between five and nine items defines the maximum capacity of working memory when full attention is deployed.

Task analysis could tell us that the TF specification is a task which puts high demands of working memory. When the subjects use a TF interface to search for required results, they continuously input different parameters of TF through the interface and judge whether the corresponding rendering results are the ones they need. Often they need retrieve previous settings that are better after comparison. The users need to perform so many interactions (the change of TF parameters and the corresponding visual feedbacks) and have to hold information in working memory, which introduces the possibility of error. The loss of information leads to unnecessary repetitive work. Clearly, the failure of human memory can have a major impact on the effectiveness and efficiency of an TF interface. It also indicates that a user interface will be more efficient and effective if it could relieve the workload of an user's working memory.

6. CONCLUSION

The described experiment is a first step towards a more quantitatively investigation of the usability of user interfaces for TF specification in volume rendering. More specifically, we have compared five interface prototypes in order to find out if specific instances of data-dependent and data-independent feedback can assist the users in this task. The obtained results can be summarized as follows:

- There is no evidence that histograms help to improve user performance.
- Additional information about possible/suggested TFs may be useful to novel users, but could also not improve performance.
- Interfaces that restrict the number of DOF of the TF also do not improve performance, and are moreover not better appreciated by users.

The trial-and-error method is a basic and important scheme to help users interact with the TF because it assigns the user as a central role in the interaction. The data-dependent and data-independent feedback mechanisms that were proposed in this paper did not substantially improve the effectiveness and efficiency of the interface. The working memory theory clearly explains the reason why subjects do failed to finish the tasks effectively and efficiently: the interfaces do not provide any mechanism to relieve the workload of working memory. This suggests that the designer of TF user interfaces should also take some human factors into account, besides the look-and-feel features of the interface. The logical next step of our research is to investigate how image-centric and automated methods, that are most useful to improve the initialization phase of the interaction, compare to the baseline interface. Future experiments, that can profit from the experimental methodology presented in this paper, should help to more quantitatively address this question.

ACKNOWLEDGEMENTS

This research is sponsored by the innovation-driven research program on Man-Machine Interaction (IOP-MMI), that is funded by SenterNovem under the Dutch Ministry of Economic Affairs.

REFERENCES

[1] A. J. Ahumada. Computational image quality metrics: A review. *Society for Information Display International Symposium, Digest of Technical Papers*, 24:305–308, 1993.

[2] A. Baddeley. *Working memory*. Clarendon, Oxford, 1986.

[3] A. Baddeley. *Human memory*. Allyn and Bacon, Boston, 1990.

[4] C. Bajaj, V. Pascucci, and D. Schikore. The contour spectrum. In *Proceedings of the 1997 IEEE Visualization Conference*, pages 167–173. IEEE, 1997.

[5] C. Botha and F. Post. New technique for transfer function specification in direct volume rendering using real-time visual feedback. In S. K. Mun, editor, *Proceedings of the SPIE International Symposium on Medical Imaging-Visualization, Image-Guided Procedures, and Display*, pages 227–234. SPIE, SPIE, 2002.

[6] K. Brodlie, L. Carpenter, J. Earnshaw, and R.A. Gallop, and R. Hubbard. *Scientific Visualization, Techniques and Applications*. Springer-Verlag, 1992.

[7] J. Brown. Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10:12–21, 1959.

[8] A. Eskicioglu and P. Fisher. Image quality measures and their performance. *IEEE Computer Graphics and Applications*, 43(12):2959–2965, 1995.

[9] G. Gediga, K.-c. Hamborg, and I. Duntsch. The isometrics usability inventory: An operationalisation of iso 9241-10. *Behaviour and Information Technology*, 18:151–164, 1999.

[10] T. He, L. Hong, A. Kaufman, and H. Pfister. Generation of transfer functions with stochastic search techniques. In *Proceedings of IEEE Visualization 1996 (VIS '96)*, pages 227–234. IEEE, IEEE CS, 1996.

[11] T. Jankun-Kelly and K.-L. Ma. A spreadsheet interface for visualization exploration. In *Proceedings IEEE Visualization 2000*, pages 69–76. IEEE, 2000.

[12] G. Kindlmann and J. W. Durkin. Semi-automatic generation of transfer functions for direct volume rendering. In *IEEE Symposium On Volume Visualization*, pages 79–86. IEEE, IEEE CS, 1998.

[13] A. Konig and E. Gruller. Mastering transfer function specification by using volumepro technology. In T. L. Kunii, editor, *Spring Conference on Computer Graphics 2001*, pages 279–286. IEEE, 2001.

[14] J. R. Lewis. Ibm computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 21(3):57–78, 1995.

[15] B. Lichtenbelt, R. Crane, and S. Naqvi. *Introduction to Volume Rendering, chapter 4*. Prentice-Hall, New Jersey, 1998.

[16] K.-L. Ma. Image graphs - a novel approach to visual data exploration. In *Proceedings IEEE Visualization 1999*, pages 81-88. IEEE, 1999.

[17] J. Marks, B. Andalman, P. Beardsley, and et al.Design galleries: A general approach to setting parameters for computer graphics and animation. In *ACM Computer Graphics (SIGGRAPH '97 Proceedings)*, pages 389-400. ACM, ACM Press / ACM SIGGRAPH, 1997.

[18] A. Melton. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and verbal Behavior*, 2:1-21, 1963.

[19] G. Miller. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63:81-97, 1956.

[20] N. Moray. *Handbook of perception and human performance*. Wiley, New York, 1986.

[21] H. Pfister, H. Lorensen, C. Bajaj, G. Kindlmann, W. Schroeder, and et al. The transfer function bake-off. *IEEE Computer Graphics and Applications*, 21(3):16-22, 2001.

[22] S. Potts and T. Muller. Transfer functions on a logarithmic scale for volume rendering. In *Graphics Interface 2004*, pages 57-63, 2004.

[23] J. Schulze, U. Wussner, S. Walz, and U. Lang. Volume rendering in a virtual environment. In *Proceedings of 5th IPTW and Eurographics Virtual Environments*, pages 187–198. Springer Verlag, 2001.

[24] F. Tzeng and E. Lum. A novel interface for higher-dimensional classification of volume data. In *Proceedings of the IEEE Visualization 2003*, pages 20-24. IEEE, 2003.

[25] D. Wickens and G. Hollands. *Enginerring Psychology and Human Performance*. Prentice-Hall, New Jersey, 1999.