KAV-DB: Towards a Framework for the Capture and Retrieval of Visualization Knowledge over the Web

Supriya Garg[1], Julia EunJu Nam[1], Kshitij Padalkar[1], Ming-Yuen Chan[2], Soren Laue[3], Waqar Saleem[3], Joachim Giesen[3], Huamin Qu[2], and Klaus Mueller[1]

1 **Computer Science, Stony Brook University**
2 **Computer Science, Hong Kong University of Science and Technology**
3 **Computer Science, Friedrich-Schiller-University Jena**

Abstract

Digital images have become a ubiquitous medium for information communication across a broad range of application domains. Along with this ubiquity has come an immense growth in the capabilities of the devices and tools used to produce and edit this imagery. New advances in algorithms and methods are made at a large pace, which makes staying on top of the learning curve difficult for general users and even experts. To accommodate this need, commercial devices and software packages typically provide a suite of presets and shortcuts for common tasks with intuitive descriptors, determined by extensive internal user studies and expert interaction but often without publishing the actual parameters and their settings. We describe an emerging framework which strives to externalize these practices into a centralized web-based community effort called KAV-DB (Knowledge-Assisted Visualization Data Bank), to allow coverage of algorithms and applications not currently driven by immediate commercial focus but of wide interest to the community of visualization researchers. The vision of KAV-DB is to provide a web service to capture, analyze, and retrieve parameter settings for visualization algorithms, given the data at hand. KAV-DB builds on a robust user study evaluation theory, called conjoint analysis, to formulate statistical models of method parameters extracted by ways of efficient user studies. We demonstrate the assessment and analysis stage our framework via two diverse example applications: relation-aware volume exploration and text annotation of color images.

Digital Object Identifier 10.4230/DFU.xxx.yyy.p

1 Introduction

The ever-growing arsenal of methods and parameters available for data visualization can be daunting to the casual user and even to domain experts. Furthermore, comprehensive expertise is often not available in a centralized venue, but distributed over sub-communities. As a means to overcome this inherent problem, efforts have begun to store visualization expertise directly with the visualization method and possibly the dataset, to then be utilized for user guidance in the data visualization, suggesting to the user both the visualization method and its best parameters for the data and task at hand. While this is certainly an immensely useful and promising development, one requirement remains - the matching of a newly acquired dataset with the appropriate segment of the library storing the expert knowledge. This requires one to detect and recognize the dataset's category at some level of granularity and then use this information as a library index.

We are currently devising a possible framework for accomplishing both stages of this process. The first stage is comprised of a data categorization, using data classification via

Dummy Event this volume is based on.
Editor: Editor; pp. 1–11
DAGSTUHL FOLLOW-UPS Dagstuhl Publishing
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Germany

a rich set of feature vectors sufficiently sensitive to detect critical variations. In a recent publication [9] we demonstrated the utility of such a framework by ways of a set of medical and computational datasets and visualized the resulting categorization as a layout in 2D. While we have built our models only by ways of k-means clustering to determine descriptive feature groupings based on similarity, this simple scheme worked already quite well. We demonstrated the capabilities for a variety of scenes of sampled objects and phenomena, consisting of medical and flow data. However, we believe that a richer set of object dichotomies could be obtained by introducing probabilistic techniques, such as Expectation Maximization (EM), into the framework, to discover more specific objects in the data. This is subject of currently ongoing work.

The second stage, more elaborated on in this article, is a framework that strives to externalize these practices into a centralized web-based community effort called KAV-DB (Knowledge-Assisted Visualization Data Bank), to allow coverage of algorithms and applications not currently driven by immediate commercial focus but of wide interest to the community of visualization researchers. The goal of KAV-DB is to advise users on the best settings for their algorithm parameters, given the data at hand. KAV-DB builds on a robust user study evaluation theory, called conjoint analysis, to formulate statistical models of method parameters extracted by ways of efficient user studies. Conjoint analysis allows one to decouple parameter effects in the user observations which enables the efficient testing of multi-parameter problem spaces.

Using web-based communities to obtain answers to scientific questions has become increasingly popular. Buchanan and Smith [3] introduced a questionnaire-based framework for personally assessments, where they found that the test results so obtained compared well with those obtained via paper and pencil tests. Internet-based surveys have also become commonplace [11], yet Sills and Song [12] mention that the low return rates with these surveys and the need for tech-savvy population. On the contrary, we have found that the internet-based survey enables one to reach many (and geographically distributed) users in a short period of time - more than one could find in a scheduled lab-based study. We suspect that if users can do the study on their own time and in a familiar environment they tend to be more willing to dedicate a few minutes for it. Proof for this phenomenon is Amazon.com's Mechanical Turk, which is able to recruit a massive amount of user study subjects in a very time. While micro-payments are required to solicit these responses, the benefits to be gained far surpass these costs. The Mechanical Turk framework was recently explored for visualization research by Kosara and Ziemkiewicz [8] and also by Heer [7].

2 Motivation and Overview

Knowledge-assisted visualization (KAV) is a new trend that seeks to augment visualization algorithms with expert knowledge, in order to make navigation through parameter spaces easier for users unfamiliar with a given algorithm, or data visualization in general. Major challenges in this undertaking are how this knowledge is collected and stored, and how it is indexed by the data. Overcoming these challenges is not purely an engineering problem - it also requires one to formalize, at least to some extent, the cognitive processes of the human observer in the loop. Although modeling human cognition as a whole remains an unsolved problem, and will be for some time, attempting to establish an understanding at a restricted scale is generally more feasible. KAV-DB follows this path by providing a framework to derive psycho-visual models/laws at the scale of individual visualization algorithms that are not derived from first principles but from measurements in web-based user studies. Here the space of algorithms can be arbitrarily fine-grained. For example, a general volume visualization algorithm may be divided by task, such as the highlighting of a certain type of data feature.

As noted above, effective user studies can produce insight into the specialized cognitive models that underlie specific visualization tasks, and this is the approach taken by KAV-DB. For a given algorithm and dataset, subjects choose, from of a variety of images produced at a diverse range of parameter settings, those images they find most favorable under the algorithm's task. This choice may be just a binary accept/reject or it may be a selection. The corresponding, favorable parameter settings are then stored in the KAV-DB database to form the knowledge for this specific algorithm and dataset. Repeating the user study for a variety of different datasets then yields the algorithm's KAV-DB knowledge base. These assessment and analysis capabilities of KAV-DB can be very useful to visualization researchers as they allow them to quickly rebut or confirm hypotheses about the algorithm and its parameters and thus gain knowledge about the effectiveness of the algorithm in the development process. We believe this is an important aspect of KAV, and it is this capability that we emphasize in this paper. Significant additional benefits in the spirit of KAV will result when the database can be queried with new datasets for which no user feedback has so far been assessed. This requires a suitable index by which this knowledge can be retrieved. More concretely, we can think of each assessed dataset as an *exemplar dataset*, that is, an instance of a specific *data category* (in practice we would use more than one instance per category in order to gain statistical robustness). Each such data category is defined by the response value of a suitable *descriptor* (or *classifier*) offering sufficient discriminative power for dataset categorization. Thus, the descriptor takes on the role of the index into the KAV-DB knowledgebase by which the stored favorable parameter settings can be retrieved for new (query) datasets. We would simply compute the descriptor for the candidate dataset and then use a suitable similarity metric, such as Euclidean distance, correlation, or the like, to match it with the closest exemplar data category to gain access to its most effective parameter settings stored in KAV-DB. Such a framework would recommend to a user the parameter settings for a given algorithm that best fit the dataset under investigation. These recommendations can then serve as a starting point for exploring the parameter space in the recommendations neighborhood to optimize the settings for the new data instance at hand. This would potentially cut down the search for useful visualization results tremendously. Essentially, the system serves as an expert, guiding the lesser experienced user into fruitful pastures of the parameter space.

In the following we describe the theoretical underpinnings of the assessment and analysis part of our framework, its implementation, and two example applications: relation-aware volume exploration and text annotation of color images. As noted above, the knowledge retrieval stage building on these previous stages is part of future work. As such our paper should be perceived as a position paper rather than one that reports on polished results.

3 Capturing Knowledge over the Web

We now describe a specific method to assess and represent psycho-visual knowledge about parameterized visualization algorithms in (web based) user studies. We refer to a parameterized visualization algorithm as a *master algorithm*, and identify it with the Cartesian product of its parameters. Here we denote the parameters as A_i, $i = 1, \ldots, n$ and assume that all the Ai are finite. The elements in Ai are called parameter levels. Hence the master algorithm is identified with the Cartesian product $A = A_1 \times \ldots \times A_n$. An *instance* of the master algorithm is a specific parameter setting, i.e., an element $a \in A$. Note that only the instances and not the master algorithm itself can be executed. Execution of an instance of the master algorithm means applying the master algorithm with the specified parameter setting on some input dataset. For example, the input in a volume rendering application is a 3D volume dataset (to be rendered), and the input in a gamut mapping application is a color image (to be printed).

The crucial idea behind our approach (see also [6, 16]) is that we identify each parameterized version of a given *master algorithm* by the image it generates. That is, an instance of the master algorithm, i.e., a specific parameter setting, is identified with the image generated using this parameter setting. This association enables us to learn effective parameter settings (and thus build the knowledge) via user studies. The master algorithm we considered in [6] was a volume rendering algorithm, and the algorithm we have analyzed in [16] was a gamut mapping algorithm. In Section 5 we will briefly present two further master algorithms and corresponding problems (such as volume rendering or gamut mapping), namely relation-aware volume exploration and text annotation of color images.

It is our aim to estimate to what extent each single parameter level contributes to the perceived quality of the output of any instance of the master algorithm. We estimate these contributions from data assessed in psycho-visual user studies using specific input datasets. That is, we obtain the user feedback on a few input datasets (in the volume rendering application [6] these were two volume datasets, and in the gamut mapping application [16] these were about 100 color images), visualized at different parameter levels. To assess the contribution of the individual parameter levels on the perceived quality of an image with respect to a given task (of course the task always depends on the given application and the list of possible tasks is much larger), e.g.,

(a) does the image allow to detect a certain structure,
(b) does the image reflect a given property, or simply
(c) is the images aesthetically pleasing?

We use *conjoint analysis*. Conjoint analysis comprises a family of psycho-physical scaling techniques, see [4]. The assessment stage of conjoint analysis involves a conjoint measurement, i.e., a measurement on an element of A (jointly on all parameter values present in the element) or more generally on an element of A_k. That is, in a conjoint measurement the parameter values are considered jointly. Here we consider two types of conjoint measurements:

(1) On an element (image) $a \in A$ a binary response of a subject (user) is measured, e.g., it is measured if the subject likes or dislikes this image, or it is measured if the subject could correctly identify a structure in the image. Note that at this stage it is not obvious how the different parameter levels present in the image a contributed to the observed outcome of the measurement. We want to refer to this type of measurement as *binary response measurement*.

(2) It is measured which out of $k \geq 2$ elements (images) $a_1, \ldots, a_k \in A$ is chosen (preferred) by a subject. Again, at this stage it is not obvious how the different parameter levels present in the images $a_1, \ldots, a_k$ contributed to the measured choice. This type of measurement is known as choice based conjoint analysis since the individual has to choose from k options. In the following we want to restrict our discussion to choice based conjoint analysis with $k = 2$ choice options. We want to refer to this type of measurement as *binary choice measurement*.

Note, that both measurements that we considered here provide binary data (labels). From these labels we seek to learn a linear value function $v : A \to \mathbb{R}$ that assigns the estimated perceived value to every element (image) in A. Linearity means that the function v can be decomposed as:

$$v(a) = v((a_1, \ldots, a_n)) = \sum_{i=1}^{n} v_i(a_i) \tag{1}$$

where $v_i : A \to \mathbb{R}$ are called *partworth* value functions, i.e., $v_i(a_i)$ is the partworth that parameter level $a_i \in A_i$ contributes to the value of an image $a = (\ldots, a_i, \ldots)$, i.e., the parameter level a_i was present in the master algorithm when the image $a \in A$ was generated. Computing the partworth value functions is a classical scaling problem and many statistical approaches have been developed over the years. Here we want to follow a quite specific approach, namely we consider the scaling problem as a binary classification problem, i.e., computing the partworth values from the binary labels provided in the conjoint measurements, see also [15]. A successful standard approach to binary classification problems are soft margin Support Vector Machines (SVM) [10] that we are also using here, both for the binary response and the binary choice measurements.

Assessing data in a conjoint study is a tedious task, especially since quite a few subjects are needed to obtain a sufficient data base (the rule of thumb says that every parameter level should be covered by at least 10-15 measurement points). Traditionally, psycho-visual studies including [6] and [16] are conducted in a controlled environment, i.e., defined lighting conditions, calibrated monitors etc. Recent results of Sprow et al. [13] demonstrate that in the context of image quality measurements web-based studies can provide comparable results to the same test conducted in a lab-based environment. Also important for our vision of KAV-DB is that the studies of Sprow et al. included conjoint measurements and proved the usefulness of the web as a platform for scaling psycho-visual studies.

4 Retrieving the Knowledge

Since we consider every input data set in a separate user study we can identify the instances of the master algorithm, i.e., the elements in $A = A_1 \times \ldots \times A_n$, with the images generated with the specific parameter setting on the data set. That is, we can consider an element $a \in A$ as an image. Then, given a similarity measure on the input data, e.g., a similarity measure on 3D volume data, or a suited similarity measure on color images, the KAV-DB can be queried using this similarity measure to retrieve similar datasets plus the corresponding user study results that then can be transferred to the query data set, e.g. as combination from the user study results for the most similar datasets stored in KAV-DB.

As mentioned, in this paper we have focused on the knowledge capture aspects of the KAV-DB, and thus our present system does not provide these retrieval capabilities at the current time. However, it is planned that KAV-DB will provide the required similarity measures and tools to transfer the user study results to new dataset.

5 Implementation and Setup of the KAV-DB Study Server

For ease of online deployment, we implemented a first version of the KAV-DB user study server using the Google Web Toolkit (GWT) [1], which uses Javascript for the design and layout of webpages while providing a Java interface to the developer. The system was deployed on an Apache Tomcat 5.5 server running on a dual core 64 bit AMD machine with 512MB RAM operating on Debian Linux version 5.0.

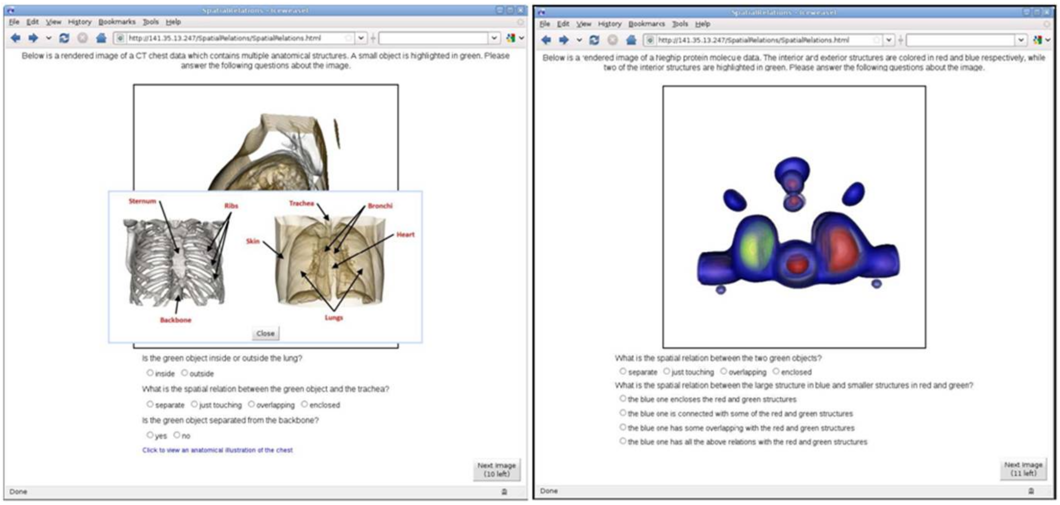

Figure 1 Sample screenshots for the volume exploration study - left: an example image from the Neghip dataset with corresponding questions, right: a helpful illustration for the questions regarding the CT chest dataset

Currently the study server hosts two studies which we will describe briefly in Section 6. Sample screenshots of each study are shown in Figures 1 and 2. Each study was designed to minimize distractions - "Are you sure?" buttons, fancy design - to keep the focus strictly on the data and the tasks at hand. A display that lists the remaining number of images in the study allows the user to stay aware of the current progress.

6 Demonstration User Studies

We tested our framework via two user studies: a 3D analysis task and a 2D color selection task.

6.1 Relation-aware volume exploration

Spatial relationships between volume structures represent crucial information in volume datasets, which is useful beyond spatial reasoning about the structures. Therefore, the visualization process should guarantee that these relationships are clearly revealed by providing views from proper viewpoints onto the volumes. However, the relation expressiveness of images can vary dramatically with the rendering parameters involved. To explore spatial information, manual or semi-automatic selection of parameters such as viewpoint, opacity, and color becomes necessary.

We first conducted a conjoint analysis user study for the *Neghip molecule* dataset which consists of layered iso-structures with various spatial relations. Figure 1 shows an example image of the data set with one particular opacity and viewpoint setting. For some settings the relationships between the structures may not be clearly shown or may even be ambiguous, while with others viewers will be able to mentally reconstruct the scene. This implies that the rendering parameters are the determinant factors to aid viewers in their spatial relation perception of the volume structures. To determine effective settings for these parameters using our framework, we rendered a set of images for the data set with different parameter settings, namely five viewpoints, two opacities, and two color schemes. Thus, in total we rendered 20 images. For a response measurement a random image was shown to a participant of our study (every respondent provided six measurement points for this study). Response measurements were performed for the following tasks:

1. What is the spatial relation between the two green objects?
   - separate / just touching / overlapping / enclosed
2. What is the spatial relation between the large structure in blue and the smaller structures in red and green?
   - the blue structure encloses the red and green structures
   - the blue structure is connected with some of the red and green structures
   - the blue structure has some overlap with the red and green structures
   - the blue structure has all of the above relations with the red and green structures

A second user study was conducted on a *CT chest* dataset in which a region was highlighted (in green), see Figure 1 for examples. For this study we also rendered 20 images for the different parameter settings (five viewpoints, two opacities, and two color schemes). Again every respondent provided six response measurement points, but this time with respect to the following three tasks:

1. Is the green object inside or outside the lung?
   - inside / outside
2. What is the spatial relation between the green object and the trachea?
   - separate / just touching / overlapping / enclosed
3. Is the green object separated from the backbone?
   - yes / no

Both datasets were used within one session of our study, i.e., collecting data from one subject. To avoid fatigue from display monotony as well as priming effects, the study alternately presented a randomly chosen image from each dataset, along with the relevant questions. A helpful illustration could be popped up by respondents for help with detailed questions. We opted to force participants to respond to all questions (the next image is not displayed until all questions for the current one have been responded to). While this increases the load on the user by forcing them to respond when unsure, it gives us valuable information in terms of which parameter settings assist them in reaching their decisions and which are unhelpful or even misleading. Thus, in this case, we opted in favor of data collection over easing user load.

6.2 Annotating color images

Annotation of an image with text is a commonly encountered task. Bauer et. al [2] have shown that it is easier to find a target color outside the $L*a*b*$ color-space convex hull of the image colors, provided the target color is linearly separable in chromaticity, or is a combination of luminance and chromaticity from the convex hull. We build on these insights and designed an algorithm that given an image and a location automatically picks a color for the annotation. For this algorithm we localized the insights of Bauer et al. by considering nested boxes around the location within the image. For every box we compute the convex hulls of the image colors within the box in the chromaticity space of $L*a*b*$ color-space. This provides us with a distance function on the chromaticity space, namely every color point is assigned the distance to the computed convex hull. Note that this distance function is zero for all points contained in the convex hull. We combine the distance functions for the nested boxes into a single function by taking a weighted sum. For our study we considered three

Figure 2 Screenshot for the text annotation study: respondents were asked to click on the image where they felt the annotation was easier to read.

boxes, namely the bounding box of the annotation, the whole image, and one box in between the first two, as well as 15 different weight vectors (with three weights each, which are used to combine the distance functions for the different boxes). Thus the master algorithm we consider here has two parameters with three and 15, respectively, levels.

We conducted choice conjoint studies for five different data sets, i.e., five different images. As with the previous studies a randomly selected pair of images from each dataset was shown one after the other in order, i.e., a pair from the first dataset followed by one from the second and so on, where the ordering of datasets was arbitrary. Respondents were asked to simply click on the image in which the annotation is better visible (Figure 2 shows an example). Then, and only then, the next image was displayed. The study ended when 15 images, three from each dataset, had been responded to.

In a conjoint study we may also consider parameters that are not parameters of the master algorithm, e.g., the location of the annotation which is part of the input to the algorithm. In our studies we exploited this and considered two additional parameters, namely the location and the patch category. The patch category was computed by clustering the image patches based on their $L*a*b*$ histograms. The similarity between two patches is defined as the histogram intersection distance [14], and affinity propagation clustering [5] was used to determine the cluster centers. A patch is assigned to a cluster only if it has a similarity of at least 1/2 to the center. The patch category is then just the cluster the patch belongs to. For our study we considered four clusters.

7 Results

Using our web framework we were able to obtain measurements from 64 subjects for the annotation studies and 32 subjects for the volume exploration studies within only one day! Such a high turnout is very hard to achieve in controlled lab studies, demonstrating that the web really can be the platform of choice to scale psycho-visual user studies. We did not record any demographic parameters, but one may assume that the subjects were either graduate students or their friends since we advertised the study mainly within these circles. We also did not record the degree of expertise of our subjects in visualization or biology. We purposely did not require any background in these fields since our aim was to test our concepts for general users. Hence, our subject sample can be assumed random with regards to these parameters.

7.1 Relation-aware volume exploration

For the study on the *Neghip molecule* data set we obtained the following (of the learned partworth value models): Question (1): 0.8625%, and Question (2): 0.641%. Accuracies were computed by ten-fold cross validation, randomly partitioning all measurements into ten folds, using always nine of the folds to compute the partworth values, and finally using these partworth values to predict the outcomes of the measurements on the remaining stratum. The cross-validation value is the percentage of correct predictions averaged over all ten strata and in our case 100 random partitions into strata.

On the *CT chest* dataset we obtained the following accuracies: Question (1): 0.6625%, Question (2): 0.7166%, and Question (3): 0.795%. To estimate the error of our computed partworth values we use a re-sampling strategy: assuming we have data points from l conjoint measurements (either binary response or binary choice). We randomly sample l out of the l data points with repetitions, i.e., some data points are left out whereas others are included more than once in the sample. On the sampled data points we compute the partworth values. We repeat this procedure t times and take the average

$$\bar{v_i}(a) = \frac{1}{t}\sum_{j=1}^{t} v_i^j(a) \tag{2}$$

where $v_i^j(a)$ is the partworth of level $a \in A_i$ computed on the j^{th} sample, and the unbiased estimator for the standard deviation

$$\sigma(v_i(a)) = \sqrt{\frac{1}{t-1}\sum_{j=1}^{t}(v_i^j(a) - \bar{v_i}(a))^2} \tag{3}$$

as our estimates for the partworth values and their errors.

The sampling approach can also be used gauge the influence of eliciting the choice measurements from a population of subjects instead of a single subject. To estimate this influence we alter the sampling strategy as follows: assume we have l/k measurements from k subjects each. We sample k subjects with repetitions and use all measurements from the sampled subjects to compute the partworth values. Again, we repeat this procedure t times and compute the mean and standard deviation of the partworth values. If the standard deviations are significantly larger than for the unbiased sampling strategy, then the population of subjects is heterogeneous with respect the measurements (and thus it makes a difference if we measure for a single subject or a population). For both studies (Neghib and CT chest, respectively) we found that the population of subjects is significantly heterogeneous.

Similarly, we can gauge the influence of the rank of a measurement in a sequence of measurements. Again, assume we have l/k measurements from k subjects each. The measurements have ranks $1, \ldots, l/k$, where the first measurement has rank 1 and the last measurement has rank l/k. Now we sample l/k measurement ranks with repetitions and use all measurements of the sampled ranks from all subjects to compute the partworth values. Again, we repeat this procedure t times and compute the mean and standard deviation of the partworth values. If the standard deviations are significantly larger than for the unbiased sampling strategy, then we observe a dependence on the question rank. Possible reasons for such a dependency can be learning effects or fatigue. We were not able to detect a significant influence of the measurement rank on the computed partworth values for both studies (Neghib and CT chest, respectively).

7.2 Annotating color images

In all five studies we found that the two additional parameters (location and a patch category) which did not form explicit parameters of the master algorithm were in fact very decisive, i.e., the partworth values of their levels were large compared to the partworth values for the other parameters (which are actually parameters of the master algorithm). This finding reveals and hints at a flaw in the design of the automatic annotation algorithm investigated. If the algorithm had been working well then it should adapt to patch category and location diminishing their influence. Since this is not the case, the proposed algorithm seems not good at picking the color for the annotation.

8 Conclusions and Future Work

We have outlined our vision for KAV-DB, a framework to derive psycho-visual models/laws at the scale of individual visualization algorithms from measurements in web based user studies. We found that utilizing the web to validate and/or optimize parameterized visualization algorithms seems very promising. Since such user studies can be easily deployed using our framework the burden to obtain user feedback is mitigated. As our color image annotation studies have demonstrated, early user feedback can help to detect design flaws in visualization algorithms (hopefully at early stages of the development). It puts the user (the visualization customer) into the design loop of the algorithm, which resembles the practice of extreme programming in software development, yielding similar benefits.

At this stage we hope to stimulate feedback from the visualization research community to make KAV-DB a useful tool used by many. We plan to offer the assessment and analysis functionality described here as a web service, which will provide incentives for participation in other user studies, as a moral (or requested) payback. We were very pleasantly surprised to receive the great number of response we did in a single day, just from the institutions of the paper authors. One reason for this high resonance presumably is that the tests themselves are fairly easy, and that the study topics were interesting and fun. To scale up the service, we aim to make user of the micro-payment system as offered by Amazons Mechanical Turk.

Another current effort is aimed at adding retrieval functionality to KAV-DB, i.e., retrieving the psycho-visual models (partworth values) for a data set and a given algorithm/task. This will then form the second purpose of KAV, allowing the system to provide recommendations for specific parameter settings that are based on previous user experiences and judgments.

Acknowledgements

We would like to thank the participants of our user studies.

References

1 http://code.google.com/webtoolkit.

2 B. Bauer, P. Jolicoeur, and W.B. Cowan. Distractor heterogeneity versus linear separability in colour visual search. *Perception*, 25(11):1281–1294, 1996.

3 T. Buchanan and J.L. Smith. Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, 90(1):125–144, 1999.

4 J.C. Falmagne. *Elements of psychophysical theory.* Oxford University Press, USA, 2002.

5 B.J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–977, 2007.

6 Joachim Giesen, Klaus Mueller, Eva Schuberth, Lujin Wang, and Peter Zolliker. Conjoint analysis to measure the perceived quality in volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 13:1664–1671, 2007.

7 J. Heer. Visualization Evaluation of the Masses, by the Masses, and for the Masses. In *BELIV Workshop*. ACM Press, 2010.

8 R. Kosara and C. Ziemkiewicz. Do Mechanical Turks Dream of Square Pie Charts? In *BELIV Workshop*. ACM Press, 2010.

9 J.E. Nam, M. Maurer, and K. Mueller. Knowledge Assisted Visualization: A high-dimensional feature clustering approach to support knowledge-assisted visualization. *Computers and Graphics*, 33(5):607–615, 2009.

10 B. Scholkopf and A.J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond.* the MIT Press, 2002.

11 M. Schonlau, R. Fricker, and M. Elliott. Conducting Research Surveys via E-mail and the Web. 2002.

12 S.J. Sills and C. Song. Innovations in survey research: An application of web-based surveys. *Social Science Computer Review*, 20(1):22, 2002.

13 I. Sprow, Z. Baranczuk, T. Stamm, and P. Zolliker. Psycho-visual methods in a web-based environment.

14 M.J. Swain and D.H. Ballard. Color indexing. *International journal of computer vision*, 7(1):11–32, 1991.

15 B. Taneva, J. Giesen, P. Zolliker, and K. Mueller. Choice based conjoint analysis: Discrete choice models vs. direct regression. In *1st ECML/PKDD-Workshop on Preference Learning (PL)*, 2008.

16 P. Zolliker, Z. Barańczuk, I. Sprow, and J. Giesen. Conjoint analysis for evaluating parameterized gamut mapping algorithms. *IEEE Transactions on Image Processing*, 19(3):758–769, 2010.