

Perception-based Visual Quality Measures

Georgia Albuquerque*
TU Braunschweig
Germany

Martin Eisemann†
TU Braunschweig
Germany

Marcus Magnor‡
TU Braunschweig
Germany

ABSTRACT

In recent years diverse quality measures to support the exploration of high-dimensional data sets have been proposed. Such measures can be very useful to rank and select information-bearing projections of very high dimensional data, when the visual exploration of all possible projections becomes unfeasible. But even though a ranking of the low dimensional projections may support the user in the visual exploration task, different measures deliver different distances between the views that do not necessarily match the expectations of human perception. As an alternative solution, we propose a perception-based approach that, similar to the existing measures, can be used to select information bearing projections of the data. Specifically, we construct a perceptual embedding for the different projections based on the data from a psychophysics study and multi-dimensional scaling. This embedding together with a ranking function is then used to estimate the value of the projections for a specific user task in a perceptual sense.

Index Terms: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—; Clustering; I.3.3 [Computer Graphics]: Picture/Image Generation;

1 INTRODUCTION

Innovative approaches in visual analytics for high-dimensional data sets have presented quality measures that can be used to automatically select promising projections of the data [13, 14]. The visual analysis of such data sets usually requires projecting the data into lower-dimensional representations. However, with the increasing amount of dimensions in scientific data sets the exhaustive analysis of all possible projections requires prohibitive time. So-called quality measures have been used in a pre-processing phase to the visual exploration. They can be effectively used to rank the possible projections of the data according to one or more user tasks and reduce the number of views to be examined by sorting them or selecting the best ones.

Usually defined with respect to an exploration task, the quality measures can be defined as ranking functions for the projections. However, an open issue of these methods is that the range and distribution of ranking values depends on the algorithm of each individual measure and it is not possible to directly compare the results of the different measures. Specifically, the possible values of the measures are often relative values and it is therefore impractical to quantify the amount of structure present in a projection for a specific user task.

In contrast, we propose a quality measure to appraise the quality of a certain projection based on human perception. By means of a psychophysics study we estimate the similarity between the projections in a perceptual space and define a measure based on these observations. Compared to the previously presented methods, our

approach has the advantage that the values assigned to the projections have a direct relation to the perceptual quality impression of human beings. In our case especially, it aims at matching the perceptual quality of visual analysts. The presented measure is very general, can be trained for a variety of exploration tasks and exploits measures derived from the human visual system to rank new unknown visualizations. Our contributions in this paper are:

- a new quality measure for projections of high-dimensional data sets based on human perception;
- the measure is generally applicable, as it can be trained for different visualizations and user tasks;
- it allows comparison of a variety of quality measures by projecting their results into our perceptual quality measure;
- we show how to train, rank and optimize our method for specific user tasks.

We present results on evaluating and ranking scatterplots for two exploration tasks: finding correlation between the dimensions and separation between classes. However, the extension to other visualization methods, e.g. parallel coordinate and pixel based displays, and other user tasks, should be straightforward. The quality estimation for scatterplots is made based on two psychophysics studies concerning each user task. The first study is used to measure similarity and dissimilarity between the visualizations. The second study estimates a ranking based on the user task. Additionally, our method allows comparison of different measures directly by mapping the acquired quality values of the visualizations into our perceptual ranking scheme.

2 BACKGROUND

There have been numerous publications that proposed methods to support the exploration of high-dimensional data sets. Asimov presented the *Grand Tour* [4] as an approach to exhaustively analyze a dataset using low-dimensional projections. The main idea was to supply the user with a complete overview of the data by generating sequences of two-dimensional projections. Nevertheless, the extensive exploration of high-dimensional datasets is usually effortful and time consuming. To address this issue different quality measures to select the best views of data sets have been proposed: From the *Projection Pursuit* [7, 9] method, over the well known Scagnostics indices [16, 18], to more recent metrics for different visualization methods and user tasks [12, 8, 13, 10, 14, 2, 5]. All those measures can be used to support the visual analysis of high-dimensional data sets, but they do not necessarily relate to the analysts opinion.

An initial study towards human perception with visual quality metrics for multidimensional data has been proposed in [15]. Tatu et al. did a user study to investigate the relationship between human perception and automatically computed measures. The authors compared three different measures from [13, 14] that can be used to estimate class separation in scatterplots and showed which measure better fits the user's opinion. Our approach is different in the sense that we propose an automatic measure based on user perception itself. Specifically, the proposed distance in the ranking function of the scatterplots is optimized to resemble the distance in a perceptual embedding for scatterplots.

*e-mail:georgia@cg.cs.tu-bs.de

†e-mail:eisemann@cg.cs.tu-bs.de

‡e-mail:magnor@cg.cs.tu-bs.de

In this work we use the analysis of paired comparisons [6] where an individual expresses a preference between two mutually distinct scatterplots. Such comparison studies have been successfully used to produce rankings [6]. Recently, Wills et al. [19] proposed a method to construct a low-dimensional perceptual embedding for bidirectional reflectance distribution functions that can be used to navigate in the space of gloss and construct new materials. The perception space is built based on a user study with paired comparisons and an extended multidimensional non-metric scaling algorithm. This multidimensional scaling algorithm copes with incomplete and inconsistent dissimilarity matrices and can therefore be trained using observations from a paired comparisons study. A detailed description of the multidimensional scaling algorithm is presented by Agarwal et al. [1]. Inspired by this, we construct a similar embedding for scatterplots that, together with a ranking function, can be used as a visual quality measure to quantify the value of the projection given a specific user task.

3 OVERVIEW

In this paper we present a perception-based quality measure to appraise the quality of visualizations. The goodness values assigned to the visualization by this novel measure are based on human observations from paired comparison studies. In particular, the distances between the measure values are optimized to be consistent with an estimated perception distance between the corresponding visualizations. It can be applied for a variety of visualizations and exploration tasks. However, for the presentation in this paper we concentrated on scatterplots and the user task of two-dimensional correlation search and class separation.

Figure 1 gives an overview of our technique. It can be divided in two main phases: a training phase where the measure is trained for a specific user task and a test phase where new visualizations can be ranked by the finished measure. The method starts with the training phase. First, a set of visualizations for the specific exploration task is chosen, and then a perceptual two-dimensional embedding P for the defined task and visualization is trained based on this set. This embedding provides a perceptually motivated similarity measure to compare different visualizations. However the embedding alone is not enough to decide when a visualization is better than another, according to the exploration task. In order to compare the quality of the visualization, we initialize a second, one-dimensional space R for ranking the visualizations. In a final stage we optimize the distances in R to resemble the distances in the perceptual embedding. This optimized ranking space R is then used to evaluate the quality of new visualizations, according to their similarity to the visualizations in the ranking space.

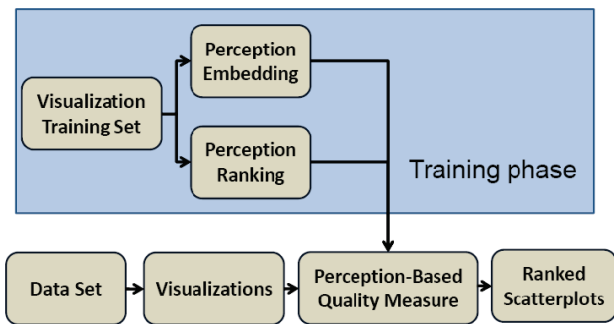


Figure 1: Working steps to select the best visualizations of a high-dimensional data set. A ranking function is defined in the training phase, based on a perceptual embedding for scatterplots. This function is used later to measure the goodness value of new scatterplots from a multivariate data set.

4 PERCEPTUAL SPACE FOR SCATTERPLOTS

Our perception-based quality measure is defined based on two user studies with paired comparisons: The first is used to estimate mutual perceptual distances between the plots (Section 4.1). The second is used to define a ranking function (Section 4.2). For each user task, a collection of scatterplots S is chosen to train the perceptual embedding and perceptual ranking. The scatterplots can be either synthetically created or taken from the projections of existing high-dimensional data sets. Choosing a representative training set S for the desired user task is of fundamental importance as it directly reflects the coverage of the final measure.

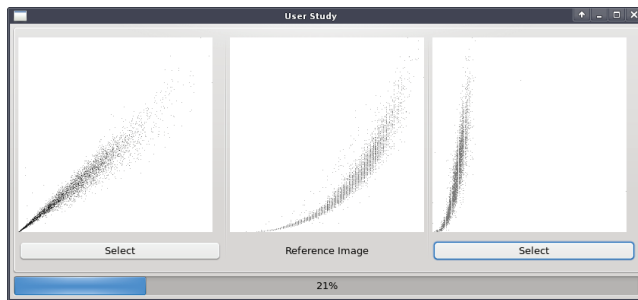


Figure 2: Screenshot of the distance comparison test. A series of scatterplot triplets is presented to each participant of the study. For each triplet, the participant is asked to decide which of the lateral images is more similar to the central one.

4.1 Perceptual Embedding

Once we have defined the exploration task, the perception embedding P for scatterplots is built in a two-step approach [19]: In the first step, a user study to estimate similarity between scatterplots is performed. In this study, a series of scatterplot triplets is presented to each participant. For each triplet, the participant is asked to decide which of the lateral images is more similar to the central one according to the specific exploration task (Figure 2). The scatterplots are randomly selected from the training set and are different from each other in the triplet. The result of the study is stored as a list of inequalities in the form:

$$L_P = \{(i, j, k) | d_{ij} \leq d_{jk}\}, \quad (1)$$

where i, j, k are indices of the scatterplots and d_{ij} denotes the perceptual distance between the scatterplots i and j . At this stage we do not know the absolute values of d_{ij} and d_{jk} , but due to the user input we know the correct ordering of i, j, k to fulfill the inequality.

In the second step, the set L_P is used to train the embedding using a general non-metric multidimensional scaling algorithm (GNMDS) [1]. Multidimensional scaling (MDS) can be defined as the process of assigning Euclidean coordinates to a set of objects based on a set of constraints. These constraints can be a set of dissimilarities, similarities or ordinal relations between the objects. The coordinates are assigned to the objects by fitting as close as possible to these constraints. The GNMDS method was developed to learn a low rank embedding from a collection of paired comparisons, applying convex optimization techniques, as described in Agarwal et al. [1]. While in the classical MDS the dissimilarity values of the objects are directly interpreted as Euclidean distances, in the GNMDS, only the relative order of the object dissimilarities is necessary. Specially, the GNMDS method has the advantage that it can be used in a variety of cases where the magnitude of the dissimilarity is uncertain or unknown, as is the case in the aforementioned paired comparison study. The method can deal with repetitions and

inconsistencies that commonly arise in such studies as distinct participants may have different opinions when comparing the scatterplots. The evaluation of a set L_P using this algorithm results in a two-dimensional embedding P of the visualizations, so that the distances in this embedding directly correlate to the user’s perception. Figure 8 shows an embedding example trained for the correlation task according to the user’s expectation of similarity between the scatterplots.

4.2 Perceptual Ranking

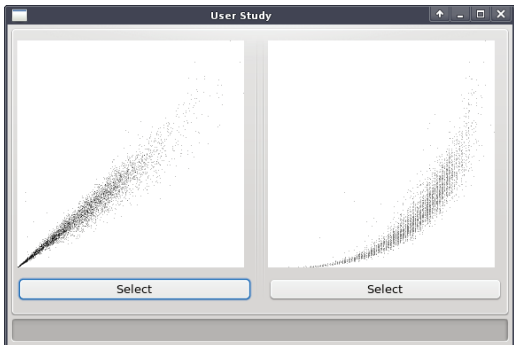


Figure 3: Screenshot of the rank comparison test. A series of scatterplot pairs is presented to each participant of the study. For each pair, the participant is asked to decide which of the visualizations has a higher quality considering a specific task.

Given the perceptual embedding P for the scatterplots, we are able to measure the perceptual distance between them. We now need an efficient way to define a ranking function for the corresponding scatterplots. Our perception ranking R is defined in a three step approach:

To estimate the ordering of the scatterplots, we perform a second paired comparison study. In this study, two scatterplots are presented to the participants and they are asked to choose the best one considering the defined user task. Figure 3 shows a screenshot of one of those tests. Similar to the first study, the presented scatterplots pairs are randomly chosen from the training set and the result is a collection of observations in the form:

$$L_R = \{(i, j) | p_i < p_j\}, \quad (2)$$

where i, j are indices of the scatterplots, p_i represents the i^{th} scatterplot and the inequality $p_i < p_j$ denotes that p_j is better than p_i , considering the chosen task.

Given the set of inequalities L_R we need an efficient way to find an optimal ordering for the scatterplots of the training set. Considering that L_R may contain inconsistencies, the problem of finding the optimal ordering for the visualizations is NP-Complete as it is equivalent to the Traveling Salesman problem. It therefore requires exhaustively searching all possible visualization arrangements. We use a greedy incremental algorithm to find a suitable order that obeys the observations in L_R as close as possible. Our approach provides a trade-off between finding the optimal solution and completing all computations in a feasible time.

We start by reducing the inconsistencies between the inequalities in L_R . We define v_{ij} as the number of observations of L_R in the form $p_i < p_j$ and \bar{v}_{ij} , the number of observations in the form $p_j < p_i$. p_i is considered smaller than p_j , only if $v_{ij} - \bar{v}_{ij} > 0$. To insert a new visualization into the current sequence we test all possible configurations for the new plot within the sequence and choose the one that best fits the observations in L_R . This is done by incrementally placing the actual scatterplot in all possible configurations of the

sequence and evaluating how the best ordering fits the observations in L_R . The algorithm is initialized with the first plot and the best placement for the second plot is computed. We then add the third plot and create all possible sequence arrangements. The process is repeated, inserting each new element one by one at the positions that fulfills the most inequalities in R until all elements have been added. The value of a sequence is defined by the number of observations that are fulfilled and is therefore defined by:

$$s = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \begin{cases} 1, & p_i < p_j \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

When all scatterplots are added, the sequence with highest value is chosen as the relative ordering for the scatterplots. Figure 4 shows an illustrative example with three visualizations. Given the set of observations $p_1 > p_2$, $p_1 > p_3$ and $p_2 < p_3$, the best placement for p_2 is found (Figure 4(c)) and the final sequence is defined (Figure 4(e)).

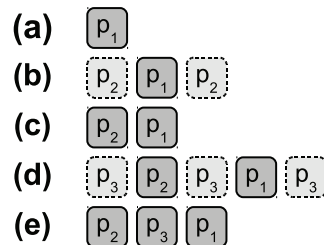


Figure 4: Placement algorithm for three example visualizations, given the set of observations $p_1 > p_2$, $p_1 > p_3$ and $p_2 < p_3$. (a) The algorithm is initialized with the first visualization p_1 . (b) All possible sequence arrangements for p_2 are tested. The set of possible insertion positions are marked with dotted boxes. (c) The best sequence is saved in next step. (d) The algorithm searches for the best placement for p_3 and the best sequence is saved at the end (e).

Up to now the ordering is only relative, in order to qualitatively evaluate the scatterplots we need to combine it with the perceptual embedding P . Our goal now is to find a ranking that, constrained to the optimized ordering, fits as close as possible the mutual distances in P . Figure 5 depicts such an example. Given an ordering of the visualization in the form $p_1 \leq p_2 \leq p_3$, and their mutual distances in the perceptual embedding d_{12} , d_{13} , and d_{23} , we want to find the best fitting goodness values for x_1, x_2, x_3 , where x_i is the rank value for the scatterplot p_i . Stating the problem in terms of a matrix $\mathbf{Ax} = \mathbf{d}$, we want to solve the following optimization problem:

$$E = \arg \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{d}\|_2^2 \quad (4)$$

$$x_i \leq x_j, i < j.$$

To solve this problem, \mathbf{x} is first initialized with the cumulative pairwise distances of the embedding P , i.e. the scatterplot ranked lowest is initialized to $x_1 = 0$ and each remaining scatterplot $x_i = x_{i-1} + d_{i,i-1}$. Afterwards, the values x_i are optimized to fit as close as possible all distances d_{ij} of the embedding P . The minimization of Equation (4) can then be computed using quadratic programming. As the problem is convex we iterate several times over \mathbf{x} , optimizing each point locally, until the algorithm converges to the global minimum.

5 PERCEPTUAL QUERY

After defining the perceptual ranking for a chosen user task, we can finally appraise the quality of new, previously unknown scatterplots. To define the goodness value of a new scatterplot p_q , we

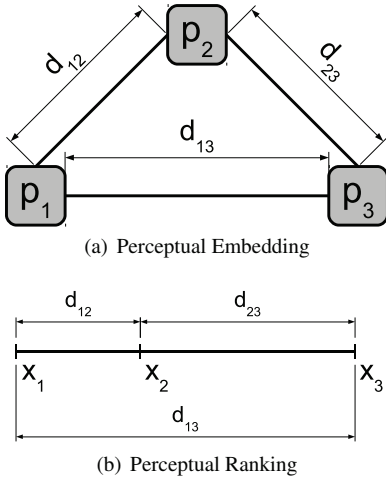


Figure 5: Embedding and ranking examples, given an ordering to the visualizations $p_1 \leq p_2 \leq p_3$, and their mutual distances in the perceptual embedding d_{12} , d_{13} , and d_{23} , we want to find the best values for x_1 , x_2 , x_3 .

search for the k -nearest scatterplots in our training set. Instead of directly comparing points of scatterplots, we choose to extract specific features to represent them and yield a more robust comparison. Principal Component Analysis (PCA) [17] is a common approach to find such robust feature descriptors of images. We follow this approach, by using the set of previously selected scatterplots in the training phase to compute an eigenobject basis. After projecting the scatterplots on this eigenobject basis, the ten first main components are chosen as feature vector.

For a new query we use a fast nearest-neighbor search [3] to find the k -best matching scatterplots in our training basis. If k is larger than 1, the influence of outliers is reduced. We use $k = 3$ throughout the paper. The final goodness value is computed as the weighted sum of the goodness values of the k best matching scatterplots. The *Perception-based Measure* (PBM) is therefore defined as:

$$PBM = \frac{\sum_{k=1}^K w_k x_k}{\sum_{k=1}^K w_k}, \quad (5)$$

where the weight w_k is defined as $\frac{1}{d_k}$, d_k is the Euclidian distance between the feature vectors of the query and the k^{th} best scatterplot. Again, it is worth noting that choosing a representative training set for the desired user task is of fundamental importance for this method because the final measure value for a new scatterplot is computed based on the values of the most similar plots in the training set.

6 EXPERIMENTS AND ANALYSIS

Our PBM measure can be trained to select high quality projections for different user tasks. For each user task, the measure is trained using a set of training visualizations to represent the task. The quality of the training set is crucial to the outcome of the PBM measure as it is always relative to the provided set. In the best case, this training set should contain evenly distributed examples of the possible space of all visualizations as the later goodness value of the visualizations is dependent on this training set. But even with incomplete training data, the results are often sufficient. We only used incomplete training data for our tests presented in this paper.

We trained our measure separately for two distinct exploration tasks: correlation finding for unclassified data and class separation for classified data, two standard tasks in visual analytics [13, 14].

For the correlation analysis task, we have chosen a set of scatterplots from the abalones data set [11], containing examples of linear and non-linear correlation, as well as different degrees of correlation. Figure 6 shows three examples of 21 scatterplots chosen to represent this task. For the class separation task, we have chosen a set of scatterplots from a synthetically generated data set [14], containing examples with two distinct classes, represented by two point clouds in two different colors. By choosing an example with only two classes, we aim to show the effectiveness of our measure for the class separation task. Note that the class separation task consists of finding scatterplots where the classes, which are already known and represented by different colors, are well separated. However, our measure can be similarly trained for the task of finding projections containing separate clusters in data sets where no label information is available. Figure 7 shows three examples of 28 scatterplots chosen to represent this task. These training sets are not complete to describe the cited tasks, but they could be successfully used to test the effectiveness of the measure.

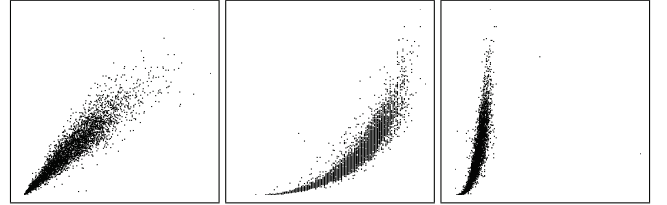


Figure 6: Examples of scatterplots from the training set for the correlation task.

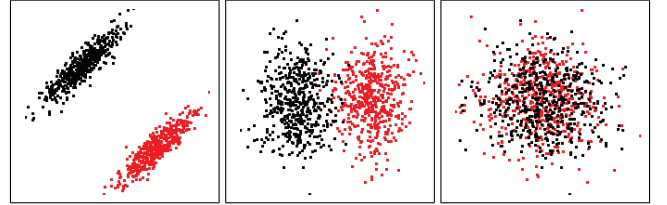


Figure 7: Examples of scatterplots from the training set for the class separation task.

For each separate task, we executed two psychophysics studies using the selected scatterplots. The participants of the studies were 20 undergraduate students, PhD students and post docs working in the field of visual computing. In the first study, 200 randomly chosen scatterplot triplets from the task training set were presented to them. They were then asked which of the lateral plots is more similar to the central one. In the second study, 100 scatterplot pairs were presented to the participants, again randomly chosen from the task training set, and they were asked to choose the best scatterplot considering the task. For the correlation task, the participants were instructed to observe the amount of correlation between the scatterplot axes and the difference between linear and non-linear correlation. Similarly, for the class separation task, they were asked to observe the separation between the class clusters, i.e. the best plots have well separated clusters and no overlap between the classes.

6.1 Perceptual Space

Using the results of the first study, we trained two perceptual embeddings, as described in Section 4.1. Figure 8 shows the resulting two-dimensional perceptual embedding trained using the scatterplots from the training set for the correlation task. It is worth noting how the scatterplots are clustered according to the *kind* (linear and non-linear) and the amount of correlation. We can clearly

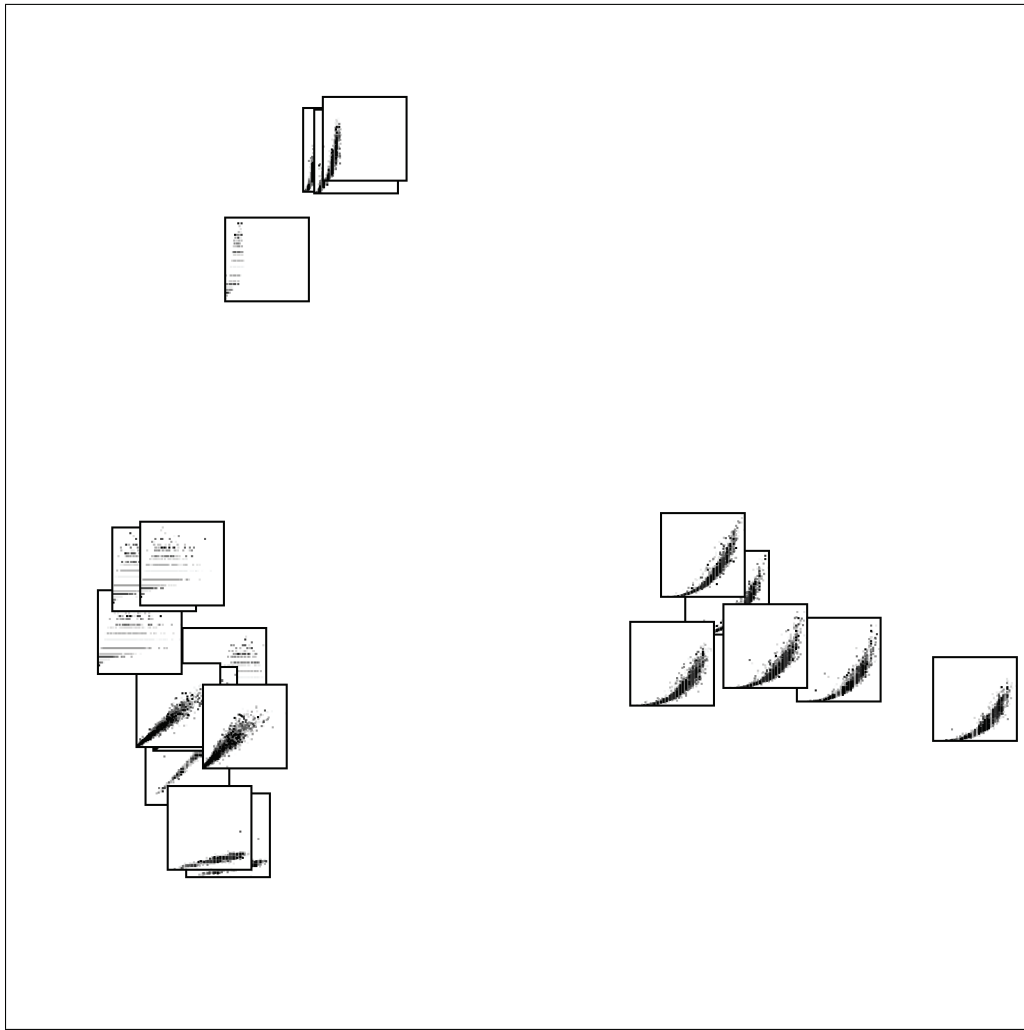


Figure 8: Resulting two-dimensional perceptual embedding trained using the scatterplots for the correlation task. The scatterplots are clustered according to the *kind* of axis correlation (linear and non-linear) and the amount of correlation. We can see three main clusters in this embedding: scatterplots presenting high and non-linear correlation cluster on the right of the embedding; lower non-linear correlation on the top and scatterplots depicting nearly linear correlation can be found on the bottom left.

see three main clusters: scatterplots presenting high and non-linear correlation clusters on the right of the embedding, lower non-linear correlation at the top, and scatterplots depicting nearly linear correlation can be found at the bottom left. Additionally, we can observe three important subclassifications in this last cluster: strong linear correlation on the middle, sparse linear correlation on the top, and lower linear correlation on the bottom.

Figure 9 shows the resulting perceptual embedding for the class separation task. Similar to the correlation space, the clusters of scatterplots can be observed according to the presented class separation. Three main clusters can be observed in this embedding as well: scatterplots with bad separation between the two classes on the right, with horizontal separation on the top-left corner and with vertical separation on the bottom-left corner.

6.2 Perceptual Ranking

The perceptual rankings for the respective tasks were created based on the second study as described in Section 4.2. Figure 10 shows the resulting ranking for the correlation task. Small-scaled scatterplots are shown in the middle row and their respective quality values are shown in the bottom row. Note that as result of the optimiza-

tion, similar plots have the same or almost the same quality value. Also this value increases according to the distance between the plots in the perceptual embedding (Figure 8). A larger discrepancy between the values can be observed when neighboring scatterplots are quite different due to the different clusters in the perception embedding. As for example in Figure 10, between scatterplots (7) and (8) where (7) behaves as an outlier and between (15) and (16) where the transition from non-linear to linear correlation occurs. The resulting order of the scatterplots resembles the observations of the second study. A characteristic that can be observed in the correlation ranking is the preference of the participants for scatterplots that present linear correlation between the dimensions. Unfortunately, some participants were unexperienced in the field of visualization and visual analytics. Therefore, these participants had problems choosing scatterplots showing the higher correlation value as they assumed correlation and skinny structures to be equal. This may result in outliers that can be minimized by using more experienced participants for the study.

Similarly, Figure 11 shows the resulting ranking for the class separation task. Again, we can observe large discrepancies between the quality values of different scatterplots. Some examples are be-

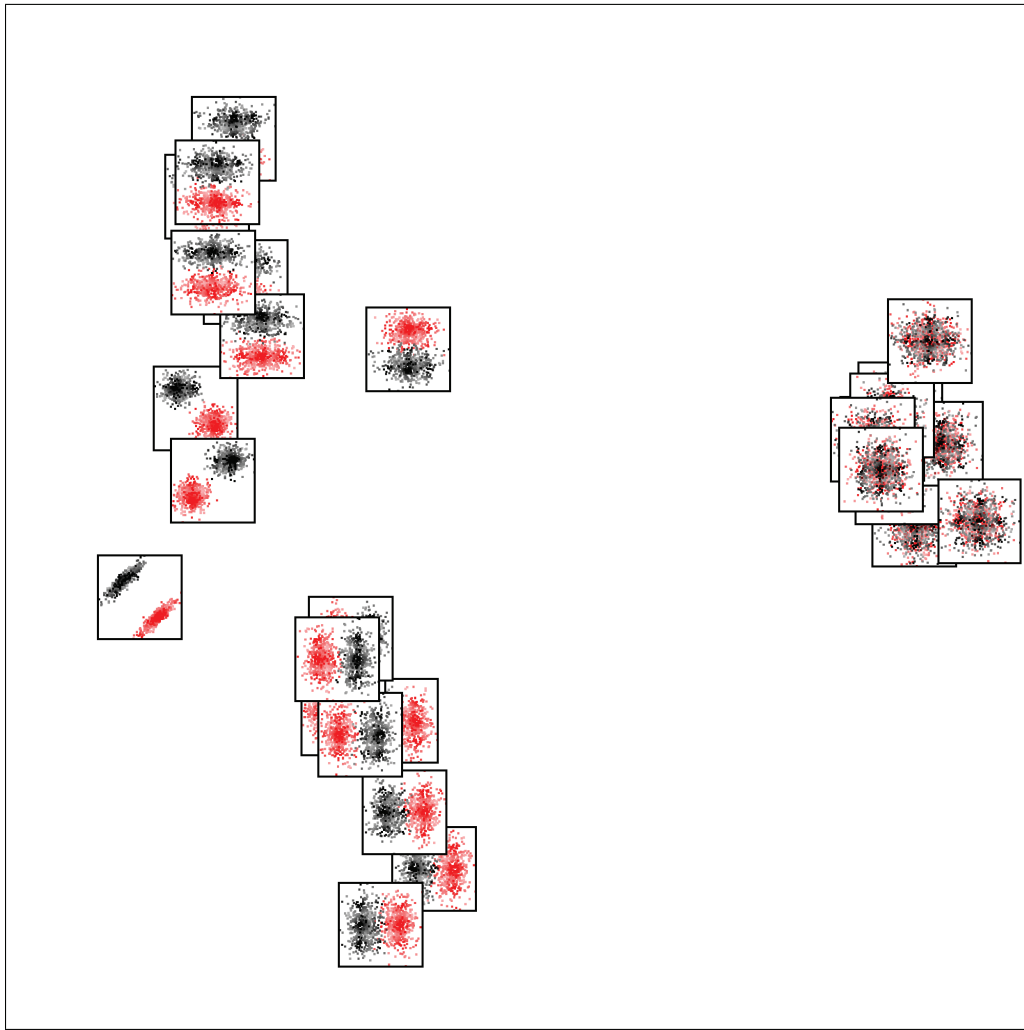


Figure 9: Perceptual embedding for the class separation task. Clusters of scatterplots can be observed according to the presented class separation. Three main clusters can be observed: scatterplots with bad separation between the two classes on the right; with horizontal separation on the top-left corner and with vertical separation on the bottom-left corner.

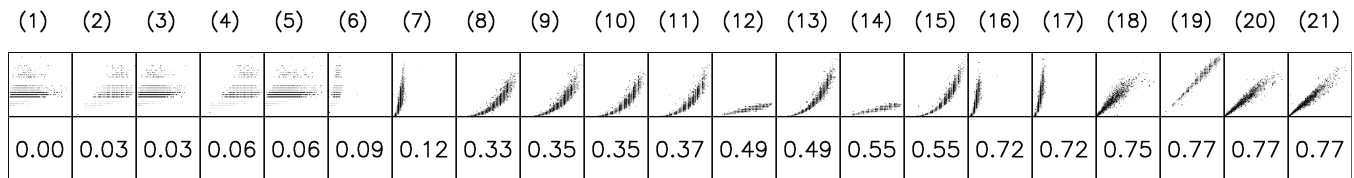


Figure 10: Perceptual ranking for the correlation task. Small-scaled scatterplots are shown in the middle row and their respective quality values are shown in the bottom row. Note that similar scatterplots have the same, or almost the same quality value and this value increases according with their mutual distances in the perceptual embedding (Figure 8).

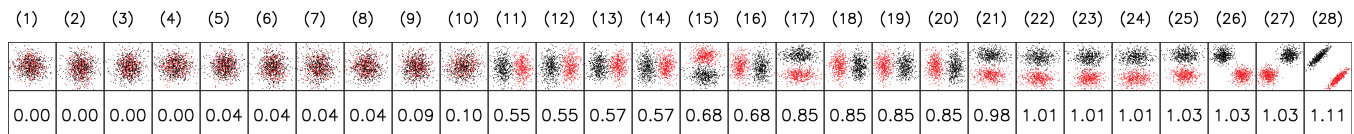


Figure 11: Perceptual ranking for the class separation task. We can observe large discrepancies between the quality values of differing scatterplots. For example between scatterplot (10) and (11), where a clear difference concerning the separation the two classes can be observed.

tween scatterplot (10) and (11) where a clear difference concerning the separation of the two classes can be observed. From (1) to (10) the scatterplots present a large overlap between the classes. An interesting aspect that can be noted in ordering of the visualizations is the preference for the horizontal separated class clusters. During the study we could observe that horizontal separated point clusters appeared to be more disjointed in the participants opinions than vertical separated ones.

To verify the stability of the scatterplot ranking, we performed a leave-one-out test with the scatterplots of the class separation task. We removed each of the 28 scatterplots and trained the measure with the 27 remaining samples. As result the newly created rankings presented an average difference of 1.28 position, compared to the original one. That means an error of 4.6%, indicating that the ranking is stable considering the dependence on the individual scatterplots of the training set.

6.3 Perception-Based Measure

To evaluate our measures we tested them with two synthetically generated data sets. To test our PBM measure with the correlation task, we used an unclassified data set with 6 dimensions and 1000 sampling points. Figure 12(a) presents the scatterplot matrix of the data set showing scatterplots above the main diagonal and their respective values according to the PBM measure under the diagonal. These values were computed based on the 3-best matching scatterplots of the training set. E.g. for the scatterplot (dim0-dim1), the scatterplots (10),(11), and (15) were automatically chosen from the training set (Figure 10) and their respective ranks were used to compute the PBM measure (Section 5). Our measure successfully ranked linear and non-linear correlation between dimensions (dim0-dim1) and (dim4, dim5) with suitable values according to the previously computed perceptual ranking. The remaining scatterplots were ranked with a low value for the PBM measure, in consistence with the perceptual ranking.

Similarly, we compute the values of the scatterplots using the RVM measure [14] (see Figure 12(b)). The RVM measure is used to find linear and non-linear correlations between pairwise dimensions. Both measures successfully select the scatterplots with the strongest correlation as best plots. Note that in the perception ranking we trained for correlation (Figure 10), the worst ranked scatterplots are similar to the scatterplots (dim0-dim2) and (dim1-dim2), what justifies their low values for the PBM. Compared to the RVM measure, the PBM has the advantage that the ranking values resemble the user perception and that it can be trained not only for correlation but for a variety of user tasks. The PBM measure has a lower sensibility due to its restrictions to the used training set, using a more comprehensive training set may increase this sensibility.

To test the PBM measure with the class separation task, we used a classified data set with 6 dimensions and 2000 sampling points and two classes. Figure 12(c) presents the scatterplot matrix of this data set showing scatterplots above the main diagonal and their respective values according to the PBM measure under the diagonal. The PBM measure successfully ranks the new scatterplots according to the previously defined perceptual ranking (Figure 11). Note that the scatterplots with horizontal class separation are rated with higher scores compared to vertical class separation due to the preference of the participants for the vertical layouts.

Figure 12(d) shows the CDM measure [14] computed for this data set. The CDM evaluates scatterplots according to the separation properties of the classes. Comparing our PBM to the CDM measure, we can observe that the PBM measure delivers more accurate values for this data set. For example the scatterplot (dim2-dim3) that presents the best separation of the class clusters is ranked with the best value by the PBM but not by the CDM measure.

7 CONCLUSION

In this paper we presented the first perceptually motivated goodness measure for qualitatively evaluating visualizations. Our new quality measure can serve several purposes. It is a very general technique that is not limited to scatterplots, but we expect it to also be applicable to other visualization methods as Parallel Coordinates or Pixel Bar Charts. We showed how to apply our technique for common visual analytics tasks such as finding projections with high (non-)linear correlations or good class separability. Further user tasks can be directly included and will be investigated in the future. We compared our method with previously proposed quality measures [15] and showed that our approach can be similarly used to rank visualizations. Our measure has the advantage that it can be trained to evaluate a variety of user tasks and that the computed quality values resemble the human perception, even though considering its limitations to the training set and user preferences.

The presented measure opens new possibilities to aid the visual exploration of high-dimensional data sets capitalizing on the human visual system. As future work, we intent to test our measure with other visualization methods and to extend the existing set of user tasks. Additionally, we aim at comparing different quality measures with different ranges in their goodness values by projecting them into our perceptual ranking function.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge funding by the German Science Foundation from project DFG MA2555/6-1 within the strategic research initiative on Scalable Visual Analytics. We also want to thank the anonymous reviewers for their many valuable suggestions.

REFERENCES

- [1] S. Agarwal, J. Wills, L. Cayton, G. Lanckriet, D. Kriegman, and S. Belongie. Generalized non-metric multidimensional scaling. In *AIS-TATS*, San Juan, Puerto Rico, 2007.
- [2] G. Albuquerque, M. Eisemann, D. J. Lehmann, H. Theisel, and M. Magnor. Improving the visual analysis of high-dimensional datasets using quality measures. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST)*, 10 2010.
- [3] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 573–582, 1994.
- [4] D. Asimov. The grand tour: a tool for viewing multidimensional data. *Journal on Scientific and Statistical Computing*, 6(1):128–143, 1985.
- [5] A. Dasgupta and R. Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 16:1017–1026, November 2010.
- [6] H. A. David. *The Method of Paired Comparisons*. Chapman and Hall, London, 2 edition, 1988.
- [7] J. H. Friedman. Exploratory projection pursuit. *Journal of the American Statistical Association*, 82(397):249–266, 1987.
- [8] D. Guo. Coordinating computational and visual approaches for interactive feature selection and multivariate clustering. *Information Visualization*, 2(4):232–246, 2003.
- [9] P. J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, 1985.
- [10] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):993–1000, 2009.
- [11] W. J. Nash and Tasmania. *The Population biology of abalone (Haliotis species) in Tasmania. 1, Blacklip abalone (H. rubra) from the north coast and the islands of Bass Strait*. Sea Fisheries Division, Dept. of Primary Industry and Fisheries, Tasmania, Hobart :, 1994.

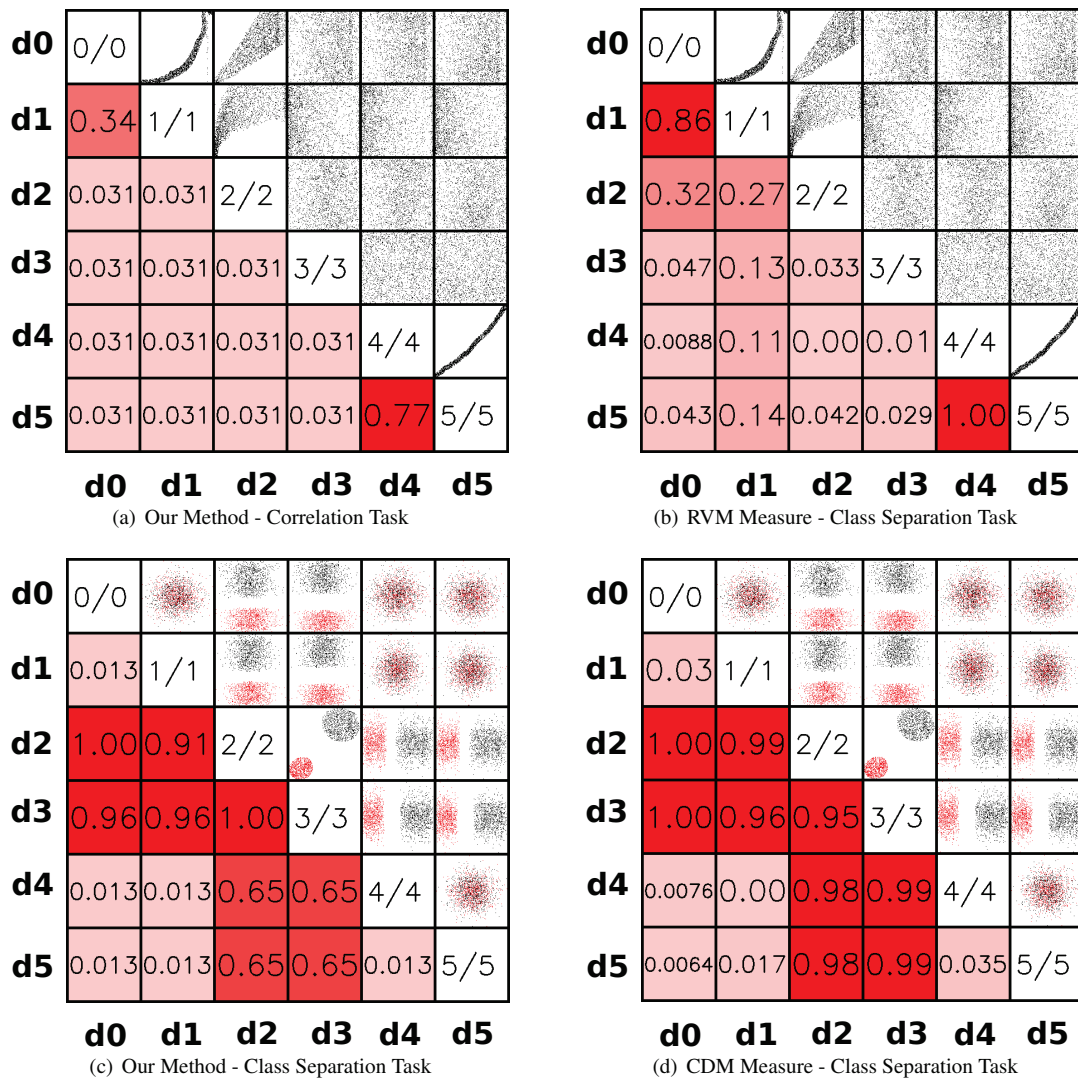


Figure 12: Scatterplot matrix of two test data sets, showing scatterplots above the main diagonal and their respective values according to the measures under the diagonal. (a) Results for the PBM measure and the correlation task, scatterplots with high correlation between the dimensions present a high quality value. (b) Results for the RVM measure. (c) Results for the PBM measure and the class separation task, scatterplots with well separated classes present a high quality value. (d) Results for the CDM measure.

[12] J. Schneidewind, M. Sips, and D. Keim. Pixnostics: Towards measuring the value of visualization. *Symposium On Visual Analytics Science And Technology*, 0:199–206, 2006.

[13] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum (Proc. EuroVis 2009)*, 28(3):831–838, 2009.

[14] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, and D. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST)*, Atlantic City, New Jersey, USA, 10 2009.

[15] A. Tatu, P. Bak, E. Bertini, D. Keim, and J. Schneidewind. Visual quality metrics and human perception: an initial study on 2d projections of large multidimensional data. In *Proceedings of the International Conference on Advanced Visual Interfaces, AVI '10*, pages 49–56, New York, NY, USA, 2010. ACM.

[16] J. Tukey and P. Tukey. Computing graphics and exploratory data analysis: An introduction. In *Proceedings of the Sixth Annual Conference and Exposition: Computer Graphics 85*. Nat. Computer Graphics Assoc., 1985.

[17] M. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[18] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 157–164, 2005.

[19] J. Wills, S. Agarwal, D. Kriegman, and S. Belongie. Toward a perceptual space for gloss. *ACM Trans. Graph.*, 28:103:1–103:15, September 2009.