Visual quality metrics and human perception: an initial study on 2D projections of large multidimensional data.

Andrada Tatu *
Institute for Computer and Information Science
University of Konstanz

Peter Bak †
Institute for Computer and Information Science
University of Konstanz

Enrico Bertini ‡
Institute for Computer and Information Science
University of Konstanz

Daniel Keim §
Institute for Computer and Information Science
University of Konstanz

Joern Schneidewind ¶
Telefonica o2 Business Intelligence Center
Muenchen

ABSTRACT

Visual quality metrics have been recently devised to automatically extract interesting visual projections out of a large number of available candidates in the exploration of high-dimensional databases. The metrics permit for instance to search within a large set of scatter plots (e.g., in a scatter plot matrix) and select the views that contain the best separation among clusters. The rationale behind these techniques is that automatic selection of "best" views is not only useful but also necessary when the number of potential projections exceeds the limit of human interpretation. While useful as a concept in general, such metrics received so far limited validation in terms of human perception. In this paper we present a perceptual study investigating the relationship between human interpretation of clusters in 2D scatter plots and the measures automatically extracted out of them. Specifically we compare a series of selected metrics and analyze how they predict human detection of clusters. A thorough discussion of results follows with reflections on their impact and directions for future research.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: [User Interfaces - Graphical User Interfaces]; I.5.3 [**Pattern Recognition**]: [Clustering - Similarity Measures]

*Email: tatu@dbvis.inf.uni-konstanz.de
†Email: bak@dbvis.inf.uni-konstanz.de
‡Email: bertini@dbvis.inf.uni-konstanz.de
§Email: keim@dbvis.inf.uni-konstanz.de
¶E-mail: schneidewind@o2.com

General Terms

User Study, Visual Quality Metrics

1. INTRODUCTION

Effective and efficient analysis of large multi-dimensional data is necessary, in order to understand the complexity of the information hidden in modern databases. Visualization has long been used as an effective tool to explore and make sense of data, especially when analysts have open-ended questions to formulate over the available information. While several techniques and commercial products have proven to be useful to provide effective support to the problem, modern databases are confronted with data complexities that go well beyond the limits of human understanding.

Data dimensionality is a major limiting factor. Finding relations, pattern, and trends over numerous dimensions is in fact difficult because the projection of n-dimensional objects over 2D spaces carries necessarily some form of information loss. Techniques like multi-dimensional scaling (MDS) and principal component analysis (PCA) offer traditional solutions by creating data embedding that try to preserve as much as possible distances in the original multi-dimensional space in the 2D projection. These techniques have however severe problems in terms of interpretation, as it is no longer possible to interpret the observed patterns in terms of the dimension of the original data space.

In order to overcome these limitations, several alternative visualization techniques have been developed in recent years, notably scatter plot matrices [3] and parallel coordinates [9], which better depict the relationship between data points and the original data dimensions. Their effectiveness, however, is highly related to the dimensionality of the data under inspection. Because the resolution available decreases as the number of data dimensions increases, it becomes very difficult, if not impossible, to explore the whole set of available projections manually.

For these reasons, a number of authors have started introducing visual quality metrics. The rationale behind this method is that quality metrics can help users reduce the search space of projections by filtering out views with low information content. In the ideal system, users can select one or more metrics and the system optimizes the visualization in a way to reflect the choice of the user.

However, one problem with these metrics is the lack of empirical validation based on user studies. These studies are in fact needed to inspect the underlying assumption that the patterns captured by these metrics correspond to the patterns that are captured by the human eye. In this paper we aim at opening a new trend of research in this direction by analyzing some of the most promising metrics.

Our analysis is based on a user study where users had to select projections of attribute-combinations well suited for classifying the data under inspection. The study then compares the scores of the selected scatter plots with the score obtained by the selected quality measures to analyze their correlation. The outcome of the study permits first of all to validate the assumption that the selection of views best ranks by quality measures is a viable way to simulate the selection of users. Furthermore, the study permits to compare the performance of the measures employed and kick-start a quality measures benchmark process, where metrics are compared against a baseline represented by the results obtained.

In summary the main contributions of this paper are:

- A validation of the hypothesis that quality measures can simulate the selection of best views by human beings
- A comparison among a set of promising and established measures
- The provision of a first benchmark framework, through which it is possible to compare new quality metrics

The rest of the paper is organized as follows. Section 2 introduces the related work, comparing our contribution to existing research results. Section 3 describes the measures employed in the study in details. Section 4 and 5 describe the whole experiment design and results respectively. Section 6 discusses the results obtained in the study offering a vision on how they can be interpreted and exploited in the future. Section 7 provides a description how to set up a framework for user based evaluation of quality metrics as suggested in this paper. Finally Section 8 provides the conclusions.

2. RELATED WORK

The two works that are mostly related to ours here are the ones from which we have selected the metrics to compare in the study ([17], [16]) which developed specifically quality measures for scatter plots. In both works the authors propose automatic analysis methods to extract potentially relevant visual structures from a set of candidate visualizations.

In [17] the visualizations are ranked in accordance with a specified user task, which corresponds to a specific metric. The ranking measures cover both classified (i.e., labeled) as well as unclassified data and can be applied to scatter plots and parallel coordinates views. From this work we include only scatter plot measures for labeled data, namely, Class Density Measure (CDM) and Histogram Density Measure (HDM).

In [16] a similar work is presented. Sips et al. provide measures for ranking scatter plots with classified and unclassified data. They propose two additional quantitative measures on class consistency: one based on the distance to the cluster centroids, and another based on the entropies of the spatial distributions of classes. The paper provides also an initial small user study where user selections are compared the outcomes of the proposed methods. From this work we adopt the Class Consistency Measure (CCM). The Class Density Measure (please note that this measure is named the same as the one used in [17] but is in fact different), which is also presented in this work, is similar to the HDM Measure and we will not include it in the analysis. Further details of these measures will be provided in Section 3.

The idea of using measures calculated over the data or over the visualization space to select interesting projections has been proposed already in some foundational works, like *Projection Pursuit* [4, 8] and *Grand Tour* [1]. Projection Pursuit searches for low-dimensional (one or two-dimensional) projections that expose interesting structures, using a "Projection Pursuit Index" which considers inter-point distances and their variation. Gran Tour adopts a more interactive approach by allowing the user to easily navigate through many viewing directions, creating a movie like presentation of the whole original space.

More recently, several works appeared in the visualization community that propose some form of quality measures. Examples are, measures based on clutter reduction for visualizations [13] [2], graph-theoretic measures for scatter plot matrices [19], measures based on class decomposition in linear projections [12], measures over pixel-based visualizations [15], and composite measures to find several data structures outliers, correlations and sub-clusters [11].

A common denominator of all these works is the total absence of user studies able to inspect the relationship between human-detected and machine-detected data patterns. While it is certainly clear how these measures can help users deal with large data spaces there are a number of open issues related to the human perception of the structures captured automatically by the suggested algorithms. In this paper we focus on the question of whether there is a correlation between what the human perceive and what the machine detects.

Despite the lack of user studies specifically focused on the issues discussed above there are a number of user studies focused on the detection of visual patterns which are worth mentioning here. A large literature exists on the detection of pre-attentive features, notably the work of Healey focused on visualization [6] and of Gestalt Laws [18] which are often taken as the basis for the detection of patterns from visual representations. Some more specific works focused on visualization are: [2] and [7] based on the perception of density in pixel-based scatter plots and in visualizations based on "pexels" (perceptual texture elements) respectively, [10] on the study of thresholds for the detection of patterns in parallel coordinates, and [5] on the correlation between the visualization performance an similarity with natural images. The study presented in [14] on feature congestion is also relevant and very similar to ours in terms of experiment design. Users ranked a series of images in terms of their perception of the degree of clutter exposed by the image and the study correlated the degree of correlation between the user rank and the rank given by the suggested *feature congestion* measure.

3. MEASURES

In the following section we will introduce the evaluated quality measures for 2D scatter plots. Our metrics come from [16] and [17] and are summarized in Table 1.

Table 1: Overview of the analyzed measures.

Measure	Section
Class Consistency (CCM)	3.1
Histogram Density 1D (1D-HDM)	3.2
Histogram Density 2D (2D-HDM)	3.2
Class Density (CDM)	3.3

In the following, the assumption is that each cluster is uniquely labeled (either manually or through some form of n-dimensional clustering algorithm) and that for each point it is possible to know to which cluster it pertains. Finally, in the visualizations shown in the paper, and those used in the experiment, each cluster is colored with a unique hue.

We will not provide extensive formal specifications and details on the metrics. For additional details and further discussions on their limits and capabilities please refer to the original papers found in [16] and [17].

3.1 Class Consistency Measure

The **Class Consistency Measure (CCM)** presented by Sips et al. in [16] is based on the distance of data points to their cluster centroid. The measure assumes the calculation of a clustering model in the n-dimensional space and computes a specific value for a given 2D projection by projecting points and centroids on the selected 2D space.

More precisely, the algorithm is based on the calculation of how many points violate the *distance to centroid measure*. For any given point the distance to its centroid in the n-dimensional space must always be lower than the distance to any other cluster centroid. But, when data is projected on a specific 2D space, this property can be violated. Therefore the measure is calculated, for a given projection, as the proportion of data points that violate the centroid distance measure.

The Class Consistency Measure (CCM) based on the centroid distance is therefore calculated as follows:

$$1 - \frac{|\{p|\exists j : d(p, centr(c_k)) \leq d(p, centr(c_j))\}|}{m} \qquad (1)$$

where c_k is the class of p, $centr(c_k)$ is the centroid of this class, m the number of available classes, and $d(p, centr(c_k))$ the centroid distance function.

3.2 Histogram Density Measure (1D and 2D)

The *Histogram Density Measure* (HDM) is a quality measure for scatter plots presented in [17]. This measure considers the class distribution of the points in the 2D scatter plot when they are projected on the axes.

In the **Histogram Density Measure 1D (1D-HDM)** data is projected over one axis and a histogram is calculated to describe the distribution of the data points over it. Since there are points pertaining to different classes (i.e., clusters) the measure is based on the analysis of the amount of overlap among points in the same histogram bin. The measure is intended to isolate plots that show good class separations, therefore HDM looks for corresponding histograms that show significant separation and this property holds when the histogram bins contain only points of one class.

In order to measure this property, the measure uses entropy and rotation. Several instances of the same 2D projection are computed, each with a different rotation factor. For each one an average entropy value is computed and the best rank among the rotation is selected as the measure's value. The computation of the entropy values is obtained as follows:

Each bin has an associated entropy equal to:

$$H(p) = -\sum_c \frac{p_c}{\sum_c p_c} log_2 \frac{p_c}{\sum_c p_c} \qquad (2)$$

where p_c is the number of data points pertaining to class C.

$H(p)$ is 0, if a bin has only points of one class, and $log_2 M$, if it contains equivalent points of all M classes.

The whole projection is ranked using the formula:

$$100 - \frac{1}{Z} \sum_x (\sum_c p_c H(p)) \qquad (3)$$

where x represents the histogram bin and $\frac{1}{Z}$ is a normalization factor, to obtain ranking values between 0 and 100.

As explained above, this is computed for every rotated projection. For each plot the best 1D-HDM output is the quality value.

The **Histogram Density Measure 2D (2D-HDM)** is an extended version of the *1D-HDM*, for which a 2-dimensional histogram on the scatter plot is computed, that is each bin represents a small square over the 2D projection and the bin count is the number of data points falling within the square. The quality is measured similarly to the *1D-HDM* by summing up a weighted sum of the entropy of each bin. The measure is normalized between 0 and 100, having 100 for the best data points visualization when each bin contains points of only one class.

In addition to the *1D-HDM*, the bin neighborhood is also taken into account in *2D-HDM*. For each bin the information of points p_c in the bin and the direct neighbors labeled as u_c are summed up. The full equation explaining the calculation in details can be found in the original paper.

The extended HDM measure to 2D can find also projections where classes are like two concentric circles of different diameters. In this case a 1D projection will always have a big overlap of the classes, even if this circles don't overlap in 2D or nD.

3.3 Class Density Measure

The **Class Density Measure (CDM)** presented in [17] evaluates the scatter plots according to their separation properties. The goal is to identify those plots that show minimal overlap between the classes.

In order to compute the overlap between the classes the method uses a continuous representation, where the points belonging to the same cluster form a separate image. For each class we have a distinct image for which a continuous and smooth density function based on local neighborhoods is calculated. For each pixel $\mathbf{p}$ the distance to its k-th nearest neighbors N_p of the same class is computed and the local density is calculated over the sphere with radius equal to the maximum distance.

Having these continuous density functions available for each class the mutual overlap can be estimated by computing the sum of the absolute difference between each pair and sum up the results.

$$CDM = \sum_{k=1}^{m-1} \sum_{l=k+1}^{m} \sum_{i=1}^{P} ||\mathbf{p}_k^i - \mathbf{p}_l^i|| \qquad (4)$$

with m being the number of density images, i.e., classes respectively, $\mathbf{p}_k^i$ is the i-th pixel in the k-th density image and P is the number of pixels. This value is large, if the densities at each pixel differ as much as possible, i.e., if one class has a high density value compared to all others. Therefore, the visualization with the fewest overlap of the classes will be given the highest value. A property of this measure is that not only it estimates well separated clusters but also clusters where density difference is noticeable, which can ease the interpretation of the data in the visualization.

4. EMPIRICAL EVALUATION

The following section describes the empirical evaluation of the described measures for projection quality. The aim of this evaluation is to assess the degree, to which these measures reflect users' perception of a high quality projection. Our method, therefore, consists of a user study for creating a baseline and a series of measures that all judge the quality of a set of scatter plots. The results show the correlation computation between all the measures with the user graded quality.

The hypotheses for the analyses were defined by the features of the four different automatic measures. We expect lowest correlation of the 1D-HDM measure with users' selection, since this measure takes only one dimensional projection for computing the separation quality of the data into account. Higher correlation results are expected by the 2D-HDM measure, because this extends its 1D version by creating a 2D histogram and considers direct neighborhoods of each data point for the quality computation. The perceived quality of a projection may be even influenced by the density of clusters having a minimal overlap, as suggested by the CDM. Finally, we expect high correlation with users' selection, when the consistency of clusters is computed, which is expressed by the quality of separation of the clusters. This is assessed by the CCM as described previously. In general, we expect a significant positive correlation of all these measure with users selection, but these measures are also expected to vary in their approximation of users' perception, which is expressed by the coefficient of determination - R^2 - of the regression.

4.1 Participants

Participants were 18 undergraduate students from the faculty of natural sciences. All had extensive experience in working with computers and scatter plots. Students participated in the experiment voluntarily and received no award for participating in the experiment.

4.2 Data and Measures

For the purpose of the empirical evaluation we took the Wine Dataset containing the results of a chemical analysis of three wine types grown in a specific area of Italy. These types are represented in the 178 samples, with the results of 13 chemical analyses recorded for each sample. This dataset is provided by the UCI Machine Learning Repository at www.archive.ics.uci.edu/ml/datasets/Wine. The 13 attributes of the dataset were pairwise combined into 78 scatter plots. The quality of these scatter plots was then computed by the four different measures. The data did not contain any special cases of cluster constellation, nor did it have outliers or hidden data points.

The number of scatter plot representations to be used in the user study was 18, in order to keep the performance time reasonably small, to allow a one-page representation of all the scatter plots at once in a reasonable size, so that all data points can be seen. The selection of the 18 scatter plots was conducted along the distribution of the measures' quality assignment, described as follows:

1. The quality values of the measures were normalized between 0 to 1, and assigned to one quantile.
2. The scatter plots were sampled in a way that the distribution between the number of projections in higher and lower quantiles is approximately the same, for all measures.
3. As a result, the distribution of quality values in each quantile was 4±1.

These selected scatter plots were ordered in six columns and three rows and printed using a high quality color printer. The order of the scatter plots was permuted by the Latin-square method, which resulted in 18 different settings, one for each participant. An example of the set of scatter plots used in the experiment is shown in Figure 1.

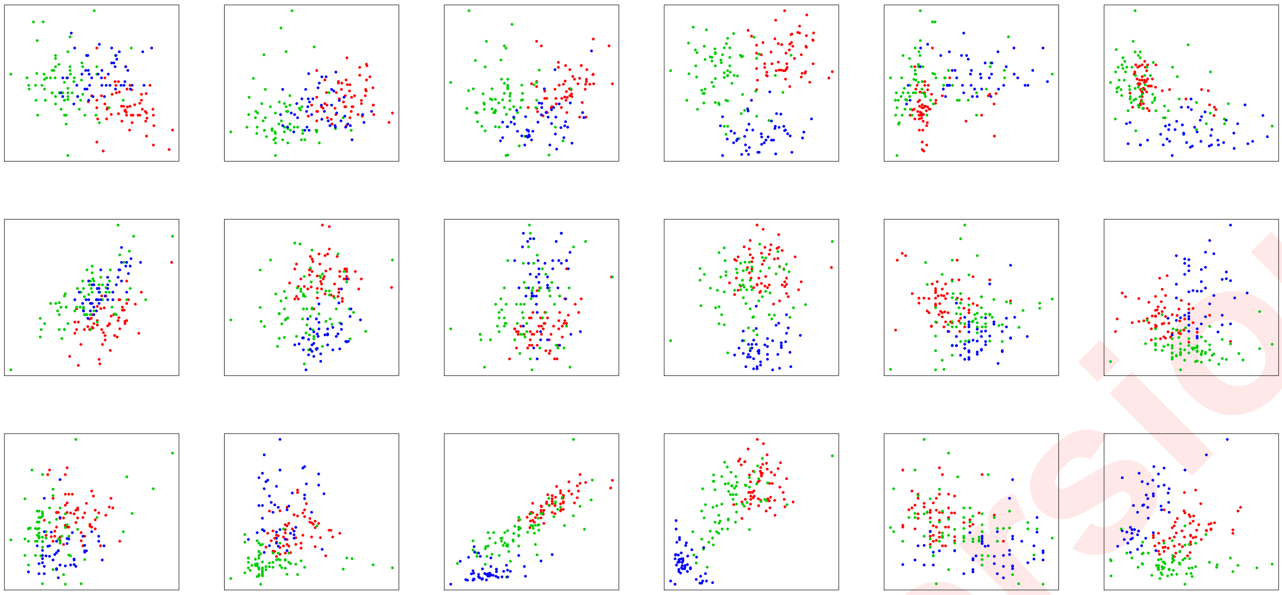


Figure 1: Projections of scatter plots used in the experiment. Participants had to select the best five projections and order them by their quality. The order of the scatter plots was permuted for each participant separately using the Latin-Square method.

4.3 Task

Participants were confronted with a scenario around the wine-dataset. They were acting in this scenario as a wine-consultants for three different types of wines. They were told that their challenge is to analyze a large amount of attributes describing the wines, such as color saturation, alcohol content, etc. Participants were requested to select projections of attribute-combinations that are well suited for classifying the three different types of wines. This task had to be carried out using a selected set of scatter plot views showing attributes in a pair-wise manner. At first, participants were asked to select the five most qualitative projections for separating wine types and then order them using numbers between 1 and 5 (5 indicating the absolute best representation, and 1 the worst out of the five best quality scatter plots).

4.4 Procedure

The experiment consisted of two parts. In the first part, participants had to read a short description of the scenario, the task and fill out a short standardized form on general questions (such as age, study stage, experience with computers and scatter plots). In the second main part of the experiment, participants had to perform the task by selecting and ordering the five best representations that classify three wine types. Clearly, the best suited scatter plot is the one that allows a clear distinction of the three wine types by the two attributes. Participants' ability to do so mainly depends on their ability to read and interpret scatter plots. The group of participants was quite homogeneous, regarding age and previous education. Expectedly, also their performance did not show significant deviations or anomalies. This was assured by computing that none is above or below the triple standard deviation. Participants were not directed on how to define a high quality projects, neither on to look for dense or consistent clusters in any way, in order not to be biased towards any of the measures.

5. RESULTS

A linear regression analysis was carried out using the Pearson coefficient for assessing the correlation between users' classification and the measures' quality assignment of the selected projections. In order to make the measures comparable, we individually normalized the assigned quality measures for the projections between 0 to 1. From the users' answers we computed the probability of selecting a projection by counting the number of times each projection was selected. These probabilities were weighted with the averaged ranks assigned by the participants. This resulted in a sequential order of the projections reflecting users' quality preferences. The dependent variable of the statistical evaluation was the user rankings, and each of the four measures was one independent variable in separate computations. Results show significant positive correlation for all four measures ($p<0.05$, DF=1, DFe=16) with the users' selection, as shown in Table 2.

Table 2: Results of the regression analysis.

Measure	t-value	StdErr.	Adj. R^2
1D-HDM	3.366	0.196	0.378
2D-HDM	6.723	0.127	0.722
CCM	6.451	0.118	0.705
CDM	5.082	0.0151	0.594

There are interesting differences in R^2 values tying the results to our hypotheses. These results indicate what proportion of the variance is explained by the regression. The highest R^2 value is achieved by the 2D-HDM, CCM performed slightly worse, followed by the CDM and the lowest by the 1D-HDM measure. Our hypotheses were partially fulfilled by these results and revealed some new significant insights. The results of the correlation are shown in Figure 2. The classification made by the users is mapped to the x-axis and by the measures to the y-axis. The charts also show the linear regression line with equation and unadjusted R^2 value.

2D-HDM and CCM assigned the best quality to the projection exactly as the users did. CDM assigned for this projection 99% quality (rank 2), and 1D-HDM only 68% quality (rank 4). The projection of users' highest quality is shown in Figure 3(a).

The highest quality projection selected by CDM and 1D-HDM is shown in Figure 3(b). This projection shows a clear and very dense cluster for one of the wine types, but a high overlap for the other two types. Users assigned rank 4 for this projection.

In users' eye the worst quality projection was the one showing high density of all three wine types but also a high overlap, as shown in Figure 3(c). This was also confirmed by three measures, except by the CDM measure, which still assigned a quality of 26.3% (rank 11) to this projection.

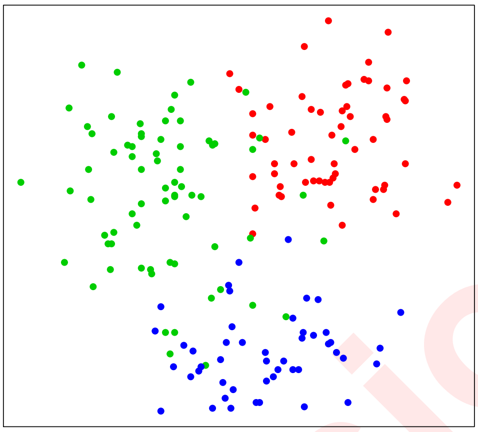
Interesting is also the phenomenon that none of the users selected 8 of the 18 projections. CDM, however, still assigned 65% quality to one of these projections. The highest quality assignment to one of these 8 projections was 58% by 1D-HDM, 50% by CCM, and only 40% by 2D-HDM.

In summary, 2D-HDM, tightly followed by CCM, reflected users' quality assignment best by getting the highest and lowest quality ranking accurately, and having the highest R^2 value of the correlation. These results should however not indicate that density (CDM) is unimportant for quality assignments, rather motivate to combine and improve these measures, so it can sufficiently support users in their task.

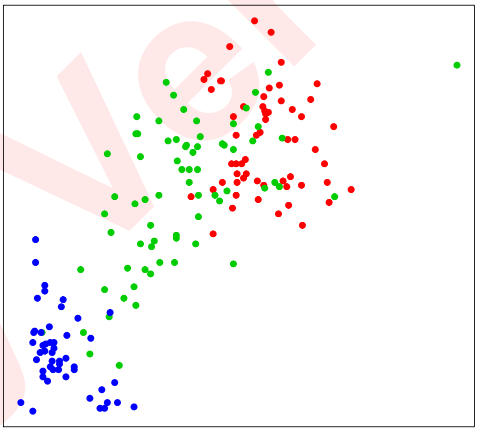
6. DISCUSSION

In the following section we examine more in details the results of the experiment, discussing some of their potential implications and ideas for further research. As we have noted in the results there is a divergence of results when the measure takes into account the density or the amount of overlap among the clusters. Histogram 2D together with Class Consistency reflected users preference for high quality projections best than the others. Intuitively, both density and and overlap should play a role in the perception of clusters, nonetheless the results of our experiment seem to suggest that separation is more important. Future research will need to address this issue and see whether a combination of measures based on both density and separation can outperform the others.

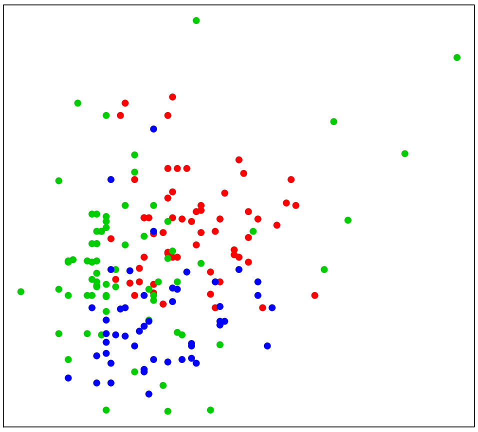
Another open issue, not investigated in this study, is the influence different shapes of clusters might have on user perception and, at the same time, on the proposed measures.



(a) Users' highest quality ranked projection was confirmed by CCM and 2D-HDM quality measures.

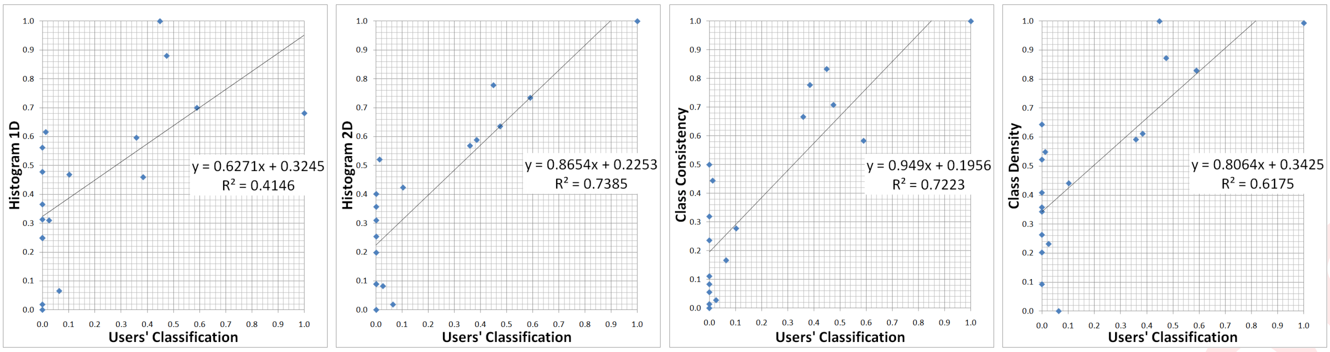


(b) Highest quality ranked projection by CDM and 1D-HDM measures.



(c) Users' lowest quality ranked projection was confirmed by CCM, 2D-HDM and also by 1D-HDM quality measures.

Figure 3: Correlation of measures with users' classification for highest and one lowest quality projection.



(a) 1D-HDM (b) 2D-HDM (c) CCM (d) CDM

Figure 2: Correlation of measures with users' classification shows highest R^2 values for the 2D-HDM measure.

Current results do not permit to differentiate between the shapes clusters have, even if the images with highly ranked clusters contain circular shapes.

In relation to this last observation, it's worth noticing that the major factor involved in the separation of clusters is the proximity of the points. This is of course not surprising as the Gestalt Laws of Grouping suggest that proximity is the strongest visual features used by the visual system to extract patterns out of images. Nonetheless, we believe it is worth running new studies investigating the relationship between the other laws of grouping (e.g., closure, similarity, continuation, etc.), users' perception and additional quality metrics.

Our experimental task is focused on the perception of clusters. However, it's important to acknowledge that the perception of clusters of an n-dimensional data spaces is not the only useful task. The detection of outliers, for instance, is also very relevant, for which it is not only necessary to find suitable metrics but also run studies similar to ours, in order to understand the relationship between user perception and the metric. The same idea can and should be repeated for several user's tasks, visual patterns, and metrics. We consider our study only a starting point in this direction, nonetheless it introduces a well-reasoned experimental design procedure that can be repeated to explore all we have outlined above. For this reason in the following section we briefly summarize the common elements of our study design so that it could be repeated in future experiments.

Finally, we want to point out that the current study focuses exclusively on the correlation and comparison of what metrics and users detect, with an underlying assumption that users' perception represents a sort of optimum. This assumption requires additional investigation as computational methods might be able to detect interesting patterns that users cannot necessarily perceive visually.

7. GUIDELINES

In the following we briefly outline the basic steps to repeat in new user studies, following the same schema used in this paper. Our motivation behind that is the desire to facilitate the design of similar studies and to promote the production of related studies on the perception of visual patterns and their formalization in computable metrics.

1. **Select a visualization technique.** The first element necessary is the selection of a specific visualization technique. In our examples we have used scatter plots which are one of the most used techniques in visualization. Future studies might include: treemaps, parallel coordinates, line charts, etc.

2. **Select a visual feature.** In this phase it is necessary to think in terms of what particular features can be detected in the visualization technique under inspection. Note that some concepts recur across several visualization but need a redefinition for each specific case (e.g., clustering in scatter plots and in parallel coordinates).

3. **Formalize the feature.** This is a fundamental step in our design schema. Once a specific features has been selected it is necessary to formalize it in a way that it can be computed through an algorithm. In this phase is advisable to produce more than one measure in order to capture several aspects of the same feature. This also permits to compare the performance of the selected in measures in the study and acquire additional information on the visual processes implied in the perception of the feature.

4. **Run a rank-based study.** Once the feature has been formalized it is possible to run a study where the users have to rank the images in terms of the selected feature. When the images have been ranked it is possible to compare the ranks given by the metrics and the ones provided by the users (as suggested in our method and design of the study).

5. **Study and refine.** The results of the algorithms can be compared to the results obtained by the users who represent the reference against which all measures are evaluated. The goal of this phase is not only to determine which of the metrics performs best, but also to reason around the results to (1)hunt for interesting insights about how users perceive the selected feature; (2)design better metrics able to capture the desired feature with more accuracy.

8. CONCLUSION AND FUTURE WORK

To conclude the research presented in this paper we would like to recall the contributions promised in the introduction. Through a user centered evaluation design we showed that some quality measures are more and some less able to reflect users' perception. However, it is still a question to which extent users are able to preselect good quality projections of their multi-dimensional data in an efficient and unbiased manner. Our results indicate that there is still a lot to be done until the ultimate automatic quality measure can be found. Nevertheless, the provision of the first quality benchmark framework, though which it is possible to compare different metrics is created. One open question regarding the future development of similar studies is whether the accumulation of several similar experiments on different visualization techniques and features can be joined to create a uniform model or better understanding of how visualization works and how visual patterns can be formalized. While the answer to this issues is not clear at the moment, it is evident that at the very least every single study has the potential to improve the understanding and the utilization of the selected technique.

In the future we plan to apply the same techniques to other visualization methods, like parallel coordinates, and therefore evaluate the correlation between the specific quality metrics and the user perception. Also we would like to investigate different visual patterns like outliers, because in the current work we focus on cluster detection exclusively. Like mentioned in Section 6, we also want to analyze how good users are in finding interesting patterns.

9. REFERENCES

[1] D. Asimov. The grand tour: a tool for viewing multidimensional data. *Journal on Scientific and Statistical Computing*, 6(1):128–143, 1985.

[2] E. Bertini and G. Santucci. Give chance a chance: modeling density to enhance scatter plot quality through random data sampling. *Information Visualization*, 5(2):95–110, 2006.

[3] D. B. Carr, R. J. Littlefield, and W. L. Nichloson. Scatterplot matrix techniques for large n. In *Proceedings of the Seventeenth Symposium on the interface of computer sciences and statistics on Computer science and statistics*, pages 297–306, New York, NY, USA, 1986. Elsevier North-Holland, Inc.

[4] J. Friedman and J. Tukey. A projection pursuit algorithm for exploratory data analysis. *Computers, IEEE Transactions on*, C-23(9):881–890, Sept. 1974.

[5] S. Haroz and K.-L. Ma. Natural visualization. In *Proceedings of Eurographics Visualization Symposium*, pages 43–50, May 2006.

[6] C. G. Healey, K. S. Booth, and J. T. Enns. High-speed visual estimation using preattentive processing. *ACM Trans. Comput.-Hum. Interact.*, 3(2):107–135, 1996.

[7] C. G. Healey and J. T. Enns. Building perceptual textures to visualize multidimensional datasets. In *VIS '98: Proceedings of the conference on Visualization '98*, pages 111–118, Los Alamitos, CA, USA, 1998. IEEE Computer Society Press.

[8] P. J. Huber. Projection pursuit. *The Annals of Statistics*, 13(2):435–475, 1985.

[9] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *VIS '90: Proceedings of the 1st conference on Visualization '90*, pages 361–378, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

[10] j. Johansson, C. Forsell, M. Lind, and M. Cooper. Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships. *Information Visualization*, 7(2):152–162, 2008.

[11] 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.

[12] Y. Koren and L. Carmel. Visualization of labeled data using linear transformations. *Information Visualization, IEEE Symposium on*, 0:16, 2003.

[13] W. Peng, M. O. Ward, and E. A. Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *INFOVIS '04: Proceedings of the IEEE Symposium on Information Visualization*, pages 89–96, Washington, DC, USA, 2004. IEEE Computer Society.

[14] R. Rosenholtz, Y. Li, J. Mansfield, and Z. Jin. Feature congestion: a measure of display clutter. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 761–770, New York, NY, USA, 2005. ACM.

[15] 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.

[16] 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.

[17] 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. *IEEE Symposium on Visual Analytics Science and Technology*, pages 59–66, 2009.

[18] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

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