

New insights into the suitability of the third dimension for visualizing multivariate/multidimensional data: a study based on loss of quality quantification

Abstract

Most visualization techniques have traditionally used 2D, instead of 3D representations to visualize multidimensional and multivariate data. In this paper, a way to demonstrate the underlying superiority of 3D, with respect to 2D representations, is proposed. Specifically, it is based on the inevitable quality degradation produced when reducing the data dimensionality. The problem is tackled from two different approaches: a visual and an analytical approach. Firstly, a set of statistical tests (point classification, distance perception and outlier identification) using the 2D and 3D visualization are carried out on a group of 40 users. The results indicate that there is an improvement in the accuracy introduced by the inclusion of a third dimension, however these results do not allow to obtain definitive conclusions on the superiority of 3D. Therefore, in order to draw further conclusions, a deeper study based on an analytical approach is proposed. The aim is to quantify the real loss of quality produced when the data is visualized in 2D and 3D spaces, in relation to the original data dimensionality, to analyze the difference between them. To achieve this, a recently proposed methodology [1] is used. The results obtained by the analytical approach reported that the loss of quality reaches significantly high values only when switching from 3D to 2D. The considerable quality degradation suffered in the 2D visualization strongly suggests the suitability of the third dimension to visualize data.

Keywords: 2D, 3D, manifold learning, dimensionality reduction, loss of quality, quality assessment criteria, multivariate data, multidimensional data, data visualization.

1. Introduction

Multidimensional Multivariate Data Visualization (MMDV) is a specific type of information visualization. This term could be adopted [2], since a set of multivariate data has a high dimensionality and can possibly be regarded as multidimensional because the key relationships between the attributes are generally unknown in advance [3, 4]. MMDV is applied to diverse areas ranging from science communities and engineering design to financial markets. Specifically in the biomedical world, MMDV is

also strongly motivated by the many situations in which expert clinicians are trying to understand the data and the inter-relationships between the massive features.

According to Keim [5, 6], MMDV techniques are classified into four categories according to the approaches taken to generate the resulting visualizations. The first, *Geometric projection*, includes techniques that aim to find informative projections and transformations of multidimensional datasets [7] such as the Scatterplot Matrix [8], the Prosection Matrix [9, 10], HyberSlice [11, 12], Hyperbox [13], Parallel Coordinates [14, 15], Andrews Curve [16], Radical Coordinates Visualization [17], Star Coordinates [18] and Table Lens [19]. The second category groups the *Pixel-oriented* techniques [7] that represent a feature value by a pixel based on a color scale. This group includes the following techniques: the Space Filling Curve [20, 21, 22], the Recursive Pattern [23], Spiral and Axes Techniques [24], the Circle Segment [25] and the Pixel Bar Chart [26]. The techniques of the third category, *Hierarchical techniques*, subdivide the data space and present sub-spaces in a hierarchical way [7]. They include: the Hierarchical Axis [27, 28, 29], Dimensional Stacking [30], Worlds Within Worlds [31] and Treemap [32]. The last category, *Iconography*, represents icon-based techniques that map the multidimensional data to different icons, or glyphs [33]. Some of them are as follows: Chernoff Faces [34], Star Glyph [35], Stick Figure [36], Shape Coding [18], Color Icon [37] and Texture [38, 39, 40].

Another way of visualizing multidimensional/multivariate data (MMD) is through the use of Dimensionality Reduction (DR), which is one of the usual operations in Data Analysis (DA) [41]. Historically, the main reasons for reducing the dimensionality of the data is to remove possible noise or redundancy in the data, and reducing the computational load in further processing. The third reason is Data Visualization (DV). One of the fields in which Dimensionality Reduction techniques for DV are currently very useful, is the scientific interactive visualization field, or Visual Analytics (VA). VA is characterised by presenting the information by making full use of the DV techniques with the possible interaction of an expert scientist. There are lot of developments in the field of VA [42] that assess it. For DV, one of the main applications of DR is to map a set of observations into a 2 or 3 dimensional space that preserves the intrinsic geometric structure of the data as much as possible [43].

Some works highlight the advantages of using 3D for DV tasks (see section 2): an additional dimension in which structures can be separated more clearly, and a reduction in the problem of overplotting. But the greatest benefit of using 3D is that a 3D view with interactive navigation controls to set the 3D viewpoint will allow users to construct a useful mental model of a dataset structure more quickly than simply by using several 2D axis-aligned views [44].

However, despite all these aforementioned advantages, one of the main drawbacks of using a third spatial dimension is strongly related to the *3D scene navigation*, since the difficulty and constraint imposed by navigating in 3D scenes have still not been overcome. There are also many other difficulties in visually encoding information with the third spatial dimension, *depth*, which has significant differences from the two planar dimensions. There are several general difficulties, such as:

1. *Line-of-sight ambiguity*: this phenomenon was defined by St. John et al [44] and describes that, we can only get information at one point along the *towards-away* depth axis for each of the rays traced from our point of view, as opposed to millions of rays that we can see along these the *sideways* and *up-down* axes by simply moving our eyes. This is because we do not really live in 3D, or even $2\frac{1}{2}$ D: in fact, we perceive the world in 2.05D [45].
2. *Occlusion hides information*: one of the most powerful depth cues is occlusion. This means that, for the main observer of the scene, a particular object may remain partially or completely hidden due to other objects located in front of it. It is possible to solve the 3D structure of the occluded elements of the scene by using an interactive navigation, but it takes time and implies a cognitive load.
3. *Perspective distortion*: this is the phenomenon in which distant objects appear smaller and change their planar position on the image plane. This distortion is one of the main dangers of depth, since the power of the plane is completely lost. For instance, if charts are used, it is more difficult to evaluate bar heights in a 3D bar than in multiple horizontally aligned 2D bars.

Furthermore, some more specific problems can be also found:

Text legibility: Another drawback derived from the use of 3D is the quality reduction in text legibility with most standard graphics packages that use current display technology [46]. Specifically, when a text label is tilted in the image plane, it often becomes blocky and jaggy.

Inappropriate view scale: If the user is placed at viewpoints too close to or too distant from the 3D scene representation, important information (e.g. 3D objects) may lie outside the viewing frustum or be so small that they go unnoticed by the user.

Limited perception of movement: Depending on the user's viewpoint and the nature of the 3D objects, it may be difficult to see objects moving toward or away from the user. For example, objects whose position or attribute change is parallel to the eye vector for the scene.

Nevertheless, these advantages and drawbacks do not univocally specify that the use of 3D is the most appropriate for MMDV. Going further, the following questions remain unanswered:

- i) When the users interact with the visualization, do they make fewer errors in 3D? Would it be possible to assess the effectiveness and efficiency when working on both representations (2D and 3D)?
- ii) Would it be possible to quantify the real degradation of quality produced in the transition from 3D to 2D? Is it too big?
- iii) Therefore, from the point of view of degradation of quality, is 3D more suitable than 2D representation to visualize data?

These open questions motivate the research presented in this paper, in which we face the challenges behind the selection of a space with the appropriate dimensionality to visualize MMD. To answer these questions, an approach focusing on *loss of quality* is proposed. This concept is the key point, since the inevitable degradation of the data quality when the dimensionality is reduced, as well as a bad choice of the data dimensionality for MMDV could drastically affect to the final interpretation of the data in the process of knowledge acquisition. Moreover, in MMDV it would be very useful to prove whether the transition from 3 to 2 dimensions generally involves a considerable loss of quality. Therefore, the final choice of 3D representations for visualization tasks would indeed be better justified.

The main hypothesis of the study presented here is based on the assertion that the use of the three dimensions in visualization counteracts the benefits of dealing with traditional 2D visualization. In other words, the intention is to demonstrate the superiority of 3D over 2D visualization for MMDV tasks.

From our point of view, the first point to be analyzed is the user's visual perception and intuition in the 3D visualization, as well as its comparison with 2D. If the results are favorable to 3D, that could shed new findings on the suitability of the use of 3 dimensions when visualizing MMD. In addition, it would be helpful to suggest to the user that there are other forms of representation (in addition to the traditional 2D) that could enhance and enrich the final visualization. So, a set of *visual statistical* tests were designed. These tests aim to highlight many of the significant differences in relation to accuracy and perception when working with a primary visualization technique such as Scatterplot, used in 3D and 2D spaces. Specifically, 3 different types of test were designed in both dimensionalities: *point classification*, *distance perception* and *outlier identification*. For each type, two measures were

designed in order to evaluate both user perception and intuition. Finally, the tests were complemented by different questions as to the suitability of visualizing MMD using 3D or 2D techniques. The tests were carried out in a random population of 40 users in an interval ranging from 19 to 75 years old. Summarizing, the results do not allow any significant conclusion to be obtained, thus it is necessary to propound the analysis problem from another point of view, the *loss of quality* produced in the transition from 3D to 2D in the DR process.

Hence, the other approach focuses on quantifying the loss of quality produced when the dimensionality of the data is reduced from 3D to 2D, which provides an analytical justification to confirm the hypothesis. This quantification is carried out using a recently methodology [1]. As far as we know, in the context of information visualization, the approximation presented here could be considered one of the first attempts at, analytically, quantifying the real loss of quality in the transition from 3D to 2D.

The rest of this paper is structured as follows: Section 2 presents a comparative analysis containing several previous studies that describe the differences, as well as highlighting the advantages and drawbacks between 2D and 3D visualization in several domains. In Section 3 the environment for carrying out the visual statistical tests is described in detail, together with a discussion of the results. Section 4 introduces an analytical approach to demonstrate the main hypothesis, and presents the results. Finally, Section 5 draws the main conclusions of the study and future lines of research.

2. Previous comparative analysis

The following studies compare 2D and 3D visualization in several domains, without focusing on the scatterplot technique in DV tasks. Therefore, many different studies have compared the visualization using only 2D and 3D views. For example, Van Orden and Broyles [47] found that 2D displays were as good as 3D displays for tasks regarding aircraft speed and altitude criteria. Park and Woldstad [48] showed that 2D and 3D visualizations were equally good for telerobotic positioning tasks. Tory et al. [49] compared 2D displays, 3D displays, and combined 2D/3D displays for relative position estimation, orientation, and volume of tasks. They demonstrated that 3D displays can be very effective for approximate navigation and relative positioning when appropriate cues, such as shadows, are present. However, 3D displays are not effective for precise navigation and positioning. Tory et al. [50] also compared point-based visualizations to 2D and 3D landscapes, where a surface has been fitted to the set of underlying points. The results showed that 2D landscapes had a better performance than 3D landscapes. In this sense, they [51] also demonstrated that the participants' visual memory was

statistically more accurate when viewing dot displays and 3D landscapes compared to 2D landscapes, and that 3D landscapes had a better performance than 2D landscapes. Smallman et al. [52] reported that 2D displays were faster when performing air control traffic tasks.

Next, several studies specifically focused on the comparison of 2D and 3D scatterplots are presented. In [53] Fabrikant compared two different kinds of display: discrete displays (a.k.a point displays) and continuous displays (a.k.a surface displays). All these displays showed dimensionally reduced data in 2D and 3D. Basically, her main contribution was to demonstrate that people could understand landscape representations of non-spatial data, as well as the relationships between 3D landscapes. She also compared point-based displays (or scatterplots), and therefore she found out that 2D scatterplots were effective mechanisms, but 3D scatterplots were more difficult to understand. Wickens [54, 55] concluded that 3D scatterplots are efficient and useful for carrying out tasks that require the integration of three dimensions. Analogously, those tasks that focused on working with one or more dimensions benefitted from a 2D scatterplot. In other words, they claimed that the *proximity compatibility principle* asserts that there is an advantage of an additional dimension when displaying the data (e.g. a 3D over two planar 2D displays, or an XY plot over two X plots) when multiple sources of data must be integrated. These conclusions indeed provide a good feedback on the advantages of 3D and 2D scatterplots. However, it is also important to note that the scatterplots that were used by Wickens et al. [56] in the experiments, showed just six or eight different points, and thus this number of points is not realistic for high-dimensional datasets, since they often contain thousands of points. Very recently, Sedlmair et al. [57] conducted an extensive empirical data study and developed a workflow model to demonstrate whether cluster separation could better performed using 2D Scatterplots, interactive 3D Scatterplots, or Scatterplot Matrices (SPLOMs). To do so, the authors analyzed a set of 816 scatterplots (derived from 75 datasets x 4 DR techniques x 3 scatterplot techniques) to assess the cluster results by using a heatmap approach. They found out that 2D scatterplots are often 'good enough', that is, neither SPLOM nor interactive 3D adds notably more cluster separability with a particular DR technique.

In contrast to previous studies, the study presented here is distinct in a number of ways:

- Firstly, a set of visual tests in 2D and 3D on a group of users are carried out. This provides the *statistical* justification.
- Secondly, in order to quantify in a numerical form the loss of quality from 3D to 2D, a methodology based on a strongly experimental section is used. This provides the *analytical* justification.

3. Visual statistical approach

Here, the environment of the visual tests carried out on a sample of users is described in detail. These tests attempt to confirm whether conclusions could be drawn as to the superiority of 3D when visualizing data, by using only the visual perception of the users.

This section is split into 3 subsections. The first one provides a complete definition of the environment needed for the carrying out of a set of tests to measure the accuracy when working with 2D and 3D visualization. The second subsection defines how the views of the users after carrying out the tests on the first part have been compiled. The last subsection presents the results of the tests, as well as a detailed discussion.

3.1. Definition of the visual tests

In order to draw valuable conclusions on the hypothetical superiority of 3D in respect to 2D when visualizing data, visualization is, indeed, required. Therefore, 3 different visual tests have been carried out on a group of users.

Motivation. The tests presented here are intended to demonstrate that the 3D visualization improves the results of the 2D visualization by using the *visual perception* of the users.

Each of the tests has been devised to yield a set of values in order to measure the accuracy (using an error value) and efficiency (using a time value) when carrying out several common tasks in DA, using 2D and 3D visualizations. These tasks are *point classification*, *distance perception* and *outlier identification* and they have been specifically designed and implemented to be used in these tests. Therefore, the values obtained when the users work in 2 and 3 dimensions for each of the three tests are compared. Finally, each user is evaluated through a set of questions that attempt to identify their personal preferences when working with 2D and 3D visualization techniques, as well as possible suggestions for the improvement of the DV.

Population sampling. To perform these tests, we were interested in sampling a set of randomly selected users from amongst the population, regardless of gender or age range, or previous experience with computers and visualization techniques. Furthermore, homogeneity in an academic background in a particular field was not a requirement.

Before starting the tests, a short series of questions were asked to the users in order to establish some basic information about them. These questions were about their gender and age range. Thus,

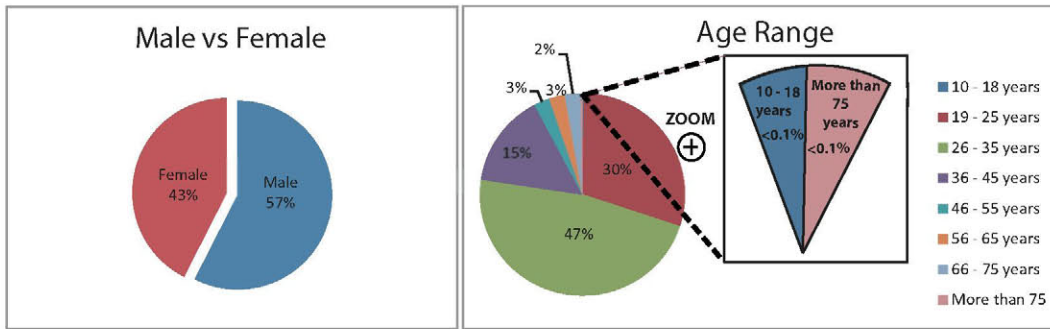


Figure 1: Basic Information on the users.

the sampling consisted of a random population of 40 users in an interval ranging from 19 to 75 years old. Figure 1 shows the summary that contains basic information on all the users who carried out the tests.

Some relevant information can be highlighted: there is a slight predominance of one gender (male, 57%) over the other (female, 43%); the age ranges most repeated are 26-35 years (47%), 19-25 years (30%) and 36-45 years (15%).

Visualization technique. Each of the 3 visual tests has been implemented for 2 and 3 dimensions. Specifically, the scatterplot technique has been selected for the visualization of the data. The rationale for using the scatterplot as the visualization technique for carrying out the tests is because it is necessary for the conclusions drawn by each user after doing the tests to be based on a simple and widely well-known visualization technique in the literature. The academical background of the test's users could be quite heterogeneous, so the selection of a visualization technique clear and understandable to all the users was an essential key point. Moreover, the representation of the data in 3 and 2 dimensions was needed, thus the scatterplot technique unequivocally provided this feature.

Data. As regards the data, one of the most well-known DNA microarray datasets in the literature has been used, Leukemia data [58], by Golub et al. (1999). This dataset was used for three main reasons. First of all, the data would have had a supervised nature, since one of the tests (point classification) requires labeled data. Secondly, the tests should be based on a highly tested and referenced datasets from other studies [59, 60, 61]. Lastly, the data should be one of the datasets used by the methodology in subsection 4.1 to quantify the loss of quality values.

DR algorithm. To represent the data selected by the scatterplot visualization technique, first it is necessary to carry out a DR process, since the data was originally of a multidimensional nature.

So, in order to successfully complete the tests, the data dimensionality is firstly reduced to 3 and 2 dimensions. Later, the 3D and 2D scatterplot technique respectively, deals with the visualization of the MMD.

The PCA algorithm has been selected for carrying out the DR. The rationale is similar to the points mentioned above, as: i) it is important to use a broadly referenced and used DR algorithm in the literature, and the PCA satisfies this requirement; ii) moreover, according the results obtained in [1], the PCA provides a high accuracy in the preservation of the intrinsic geometric structure of the data [43], which ensures a good quality in the final data visualization. Note that the PCA is of an unsupervised nature [62, 63] and the leukemia data are supervised, thus to reduce the data the original classes have not been taken into account. Subsequently, the data are coloured in accordance with their labels once they are visualized.

It is also worth highlighting that the selection of the DR algorithm could be a decisive issue for achieving different results. However, the aim of the tests is to show the perception skills, experience and criteria of the users when working with 3D and 2D data, obtained by the same method (PCA). Therefore, in this case, the user is abstracted from this particularity and is presented by a visualization of the data from the previous dataset.

Time measurement. Before explaining the details of the tests, it is important to highlight that the *time* (in seconds) that each user takes to complete each test is measured. However, the users were not notified that the time is going to be taken into account, so that nobody modifies their rhythm of work, and thus taking the needed time to properly complete each test. This will provide a better appreciation of the real time that each user takes to complete a test, by using either a 2D or 3D scatterplot. Note that the time the users had to perform the tests was open. As they finished the tests, the time taken was recorded.

Validation of the results. *Cross-validation* is used in order to properly validate the results. For each user, we followed a specific methodology (see fig.2).

It is very important to highlight three details about this methodology. Firstly, the order to carry out each of the three test has been completely serialized and randomized, that is, an user might carry out the distance perception test in the first place, while another user could perform first the point classification test. Secondly, when performing a particular test, the order in which the 2D and 3D version is presented to the user has been randomized. Finally, the random points selected to conduct each version of the test have been also randomized, and they are different for both versions of the test.

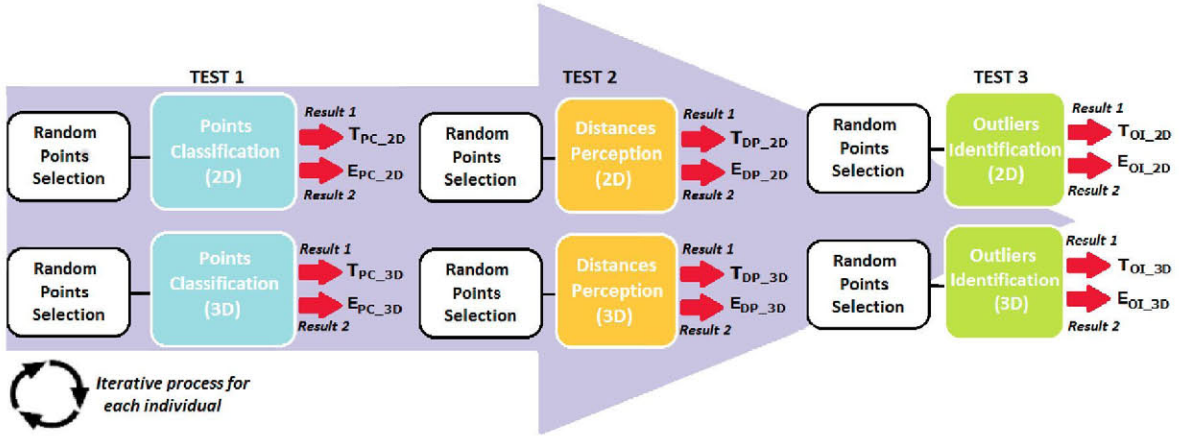


Figure 2: Methodology used in the visual tests. For each user, three different tests have been implemented. These have been presented to the users by using serialization and randomization (this means that the order used to carry out the tests could be different for the users). Furthermore, the order of presentation of the 2D and 3D version of the tests to the users is random, and the points selected to each version have been also randomized. By repeating this process for each user, a *cross-validation* of the results is achieved.

Then, and taking into account this configuration, when the users carry out the 2D *point classification* test they obtain two different results, T_{PC_2D} (time) and E_{PC_2D} (error value). For the 3D version other two results are obtained, T_{PC_3D} (time) and E_{PC_3D} (error value).

When the users carry out the 2D *distance perception* test they obtain two different results, T_{DP_2D} and E_{DP_2D} . For the 3D version other two results are obtained, T_{DP_3D} and E_{DP_3D} .

Finally, when the users carry out the 2D *outlier identification* test they obtain two different results, T_{OI_2D} and E_{OI_2D} . For the 3D version two results are also obtained: T_{OI_3D} and E_{OI_3D} .

It is worth mentioning the following points:

- The order to carry out each of the 3 tests has been randomized.
- The order to carry out the 2D and 3D version of a particular test has been randomized.
- The points selected at the beginning of each test were different for the 2D and 3D version of the test. Thus, by selecting random points for each user that carries out the test, a *cross-validation* of the results is achieved.
- All the users were shown both the 2D and 3D version of the tests.
- All the users completed all the set of tests.
- The correct solution to the tests was not shown to the users. As the users finished each test, the next test was shown.

Scene navigation. During the carrying out of the visual tests, it is essential for the user to be able to move and navigate properly through the 2D and 3D scenarios, respectively. For each scenario, either 2D or 3D, a set of controls that allow this interaction are provided.

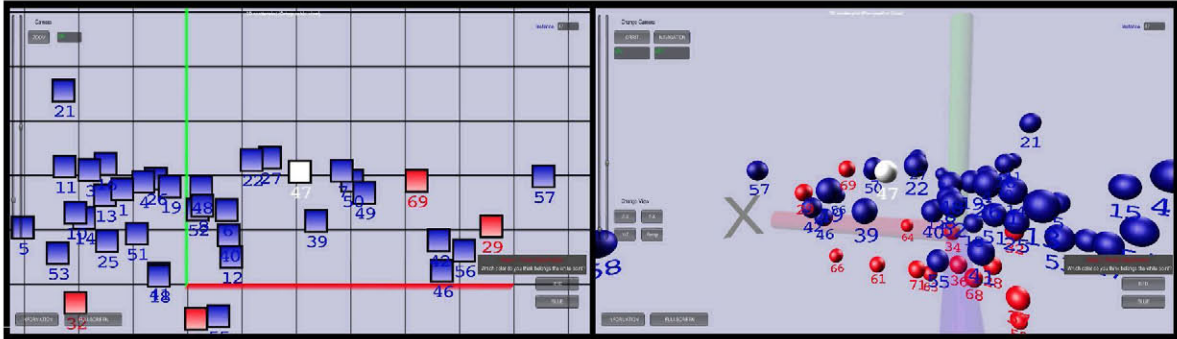


Figure 3: 2D and 3D scenarios. Each scenario provides the user different views and camera modes, as well as several sliders for adjusting the DV.

- *2D scenario.* In this scenario, a 2D ortographic view has been used. The option of scrolling vertically and horizontally is provided. Smooth and efficient zooming in and out are also provided, by using the mouse scroll wheel. Lastly, the user can conduct an automatic and smooth zoom in on particular points. This allows the view to be automatically moved and focused on those points of interest for the user (for instance, this is useful when the user is classifying points or calculating distances, see fig. 3, left-hand image).
- *3D scenario.* Two different main camera modes have been implemented: *Orbit* and *Navigate*. When using the first one, a point can be selected in order to rotate the camera around that point, by moving the mouse. The second mode allows navigation through the 3D scene using the keyboard (moving forward, moving backward, move left and move right), and the mouse (for spinning the camera). *Pan* (horizontal movement, left and right) and *pedestal* (vertical movement, up and down) camera movements have been also implemented by clicking the scroll wheel button. There is also the possibility of using a third *static* camera mode, that shows the tri-dimensional scene from different planes: Z-X, Y-X, Y-Z or perspective. By default, the camera is presented in perspective. There are some cases in which using the different perspectives, derived by a third dimension, could make the performance of a specific task easier, such as the outlier identification or distance perception (see fig. 3, right-hand image).

Two sliders have also been included for adjusting the point size, and the scale of the point position. Both adjustments are achieved by multiplying the default values by the value provided by each slider. Therefore, the user could adjust the display in order to feel and work more comfortable. The tests have been implemented by using the Unity3D visualization engine [64], which has been previously used for data visualization purposes [65, 66, 67, 68, 69, 70, 71]. Finally, before starting the tests, the importance of carefully reading the navigation controls was also emphasized.

Dissemination of the tests. To distribute the tests, the Unity3D web feature has been used. This visualization engine allows the previously developed applications to be built in web format. A web link containing the previously uploaded application was sent to each of the users who were to perform the tests. Therefore, the users performed the study via the Internet.

Others. Before carrying out of the tests, the users were provided with a detailed description (and in some cases definitions) of the working of the test, accompanied by some pictures showing the test to be carried out. The aim was to completely clarify the task before doing it, so that there was no possible ambiguity.

Furthermore, the users could test and learn the system before carrying out the real tests. Specifically, they were allowed to test the navigation controls as well as the different views to get familiarized with the interaction before performing each test.

3.1.1. Point classification

The first test is for the user to classify a set of points, which will be shown unlabeled (white). The idea is that the user says, in his opinion, whether he thinks that the point to be classified belongs to one class or another. As mentioned above, the data being displayed are related to Leukemia. Red represents ALL Leukemia (acute lymphoblastic leukemia) and blue AML Leukemia (acute myeloid leukemia). A criterion to determine whether a white point belongs to one kind of leukemia, could be based on their closeness or proximity to the blue or red group of points (see fig.4).

Motivation. The aim is to evaluate the effectiveness when carrying out the task of classifying points in a visual way, using a 2D and 3D scatterplot (in order to represent 2D and 3D MMD, respectively). For each user that carries out the test, two different numerical values are obtained, $T_{PC_{2D}}$ and $E_{PC_{2D}}$. $T_{PC_{2D}}$ represents the time taken to complete the task, and $E_{PC_{2D}}$ means the % of points that the user has successfully classified (using the 3D scatterplot technique, the obtained values will be $T_{PC_{3D}}$

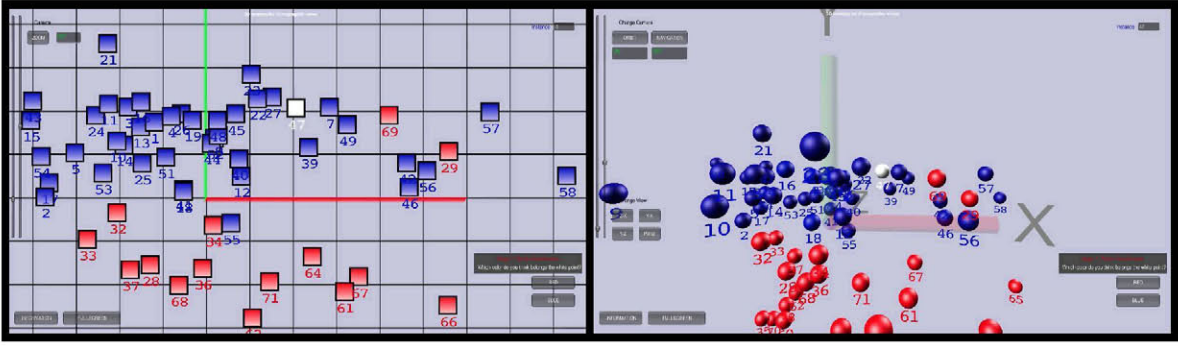


Figure 4: Point Classification test. Left-hand image: the 2D version. Right-hand image: the 3D version. The point to be classified is colored white.

and $E_{PC,3D}$). Note that, the number of correctly classified points were computed by using the original labels of the data.

Details. In this test, 10 points (from the 72 original points in the dataset) are randomly selected and removed from the visualization. These are the points that the user has to classify. Each point (white) was consecutively presented and visualized and the user was asked to say which color group the point belonged to. Once the point is classified by the user, regardless of his answer, that point was colored with its real color (according its original label). Otherwise, the user would use misinformation when classifying the following point. Finally, each label assigned a point by the user is compared with the original label of that point to obtain a value representing the number of correctly classified points.

For each user, $E_{PC,2D}$ and $E_{PC,3D}$ are computed as follows:

$$E_{PC,2D} = \frac{WC_{2D}}{T_P} 100; \quad (1)$$

$$E_{PC,3D} = \frac{WC_{3D}}{T_P} 100; \quad (2)$$

where $E_{PC,2D}$ and $E_{PC,3D}$ are the % well classified of points in the 2D/3D test, respectively. Thus, WC_{2D} and WC_{3D} represents the number of well classified points in the 2D/3D test, respectively. T_P means the total number of points to be classified and it has been set to 10. Note that the greater $E_{PC,2D}$ and $E_{PC,3D}$ values, the better accuracy.

3.1.2. Distance perception

In the second test, the user must calculate the size relationship between two lines of different color and length. Thus, two lines will be shown, yellow and magenta. The user must calculate about how big or small the yellow line is in relation to the magenta line (see fig.5).

That is, if the user thinks that the yellow line is *longer* than the magenta line, for example twice the length, the value he should say is 2. Thus, if the yellow line is equal to or longer than the magenta line, the value should be equal to or greater than 1, respectively.

Conversely, if the user thinks that the yellow line is *shorter* than the magenta line, he should give a value between [0,1]. For instance, if the yellow line is half that of the magenta line, the value should be 0.5. Finally, if the yellow line is very small compared to the magenta line, the value could be, for example, 0.2.

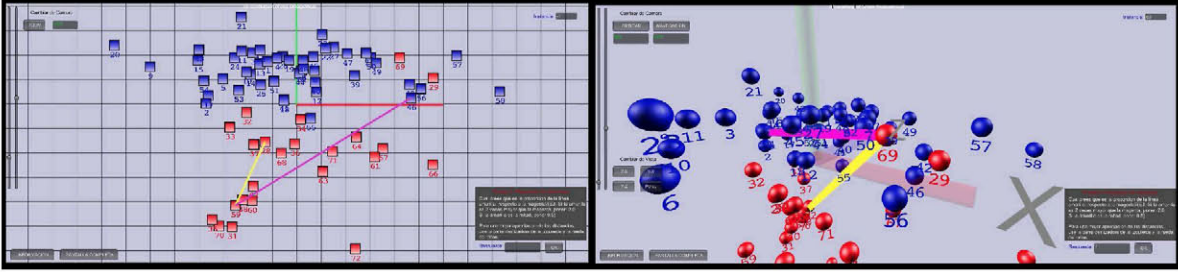


Figure 5: Distance Perception test. Left-hand image: 2D version, here the yellow line could be perceived as roughly twice the length of the magenta line, thus the value to be introduced should be approximately 2.0. Right-hand image: 3D version. Here, the inclusion of an extra dimension could provide new information about the relation, in terms of distances, between both lines.

Motivation. This test attempts to evaluate the error that an user makes when perceiving distances between points, in 2D and 3D spaces. For each user carrying out the test, two numerical values are obtained, $T_{DP,2D}$ and $E_{DP,2D}$. $T_{DP,2D}$ represents the time taken to complete the test, and $E_{DP,2D}$ is the error made by the user when calculating the distances between points. This error is computed based on the euclidean distance matrix (δ) of the original data.

Details. The following steps are taken to obtain the $E_{DP,2D}$ and $E_{DP,3D}$ errors made by the user:

1. The euclidean distance matrix of the original data is computed (without reducing the dimensionality), δ . The distance between two points, i and j , is represented by δ_{ij} .
2. Theoretically, to represent two different lines, 4 different points are needed (since a line is represented by connecting two points). To make it easier for the user to make the comparison

between these two lines, they will share a common point. Therefore, to generate two lines 3 different points are needed, instead of 4 (that is, among those three points, there is one which is connected to the other two by two different lines). Thus, those 3 points are randomly selected, named i , j and z . The yellow line will be represented by the line connecting the points i - j , and the magenta line will be represented by the line connecting the points j - z .

3. The proportion of original distances between the points i , j and z (that is, the ratio of distances calculated on δ) is defined as: $P_{ijz} = \delta_{ij}/\delta_{jz}$. This value represent the real ratio, in terms of distance, between the pair of points i - j in relation to j - z , computed on δ matrix.
4. The same relationship computed on the reduced data, that are being visualized in the test, will be named P'_{ijz} .
5. Therefore, the user should estimate the value of P'_{ijz} by visually observing the relationship between the yellow and magenta line in the display.
6. The error made by that trio of points is defined as the subtraction between P_{ijz} (real ratio) and P'_{ijz} (ratio estimated by the user). That is, the closer the value of P'_{ijz} to P_{ijz} , the smaller the error.

This process is repeated M times, starting from the step 2. In this case, a value of $M=10$ has been set. Thus, at the end of the test, the user should have evaluated M randomly selected different trios of points. It is worth mentioning that for each value of M , the selected points i , j y z will be different.

Therefore, the total error made by each user during the test, and after evaluating all the trios of points, is defined as:

$$E_{DP.2D} = \sum_{m=1}^M (P_{ijz} - P'_{ijz})^2 \quad (3)$$

where $E_{DP.2D} \in [0, +\infty]$. A value close to 0 indicates that the error is low, that is, by using the 2D visualization the user has effectively perceived and consequently estimated the real ratio between the distances in the original data. However, a value that tends toward infinity indicates that the perception of the user, in relation to those distances in the 2D space, is completely erroneous regarding the real ratio between those distances.

The process to obtain $E_{DP.3D}$ is exactly the same as explained for $E_{DP.2D}$, but using the results obtained in the 3D version of the test.

Thus, now the motivation of this test could be formally rewritten to compare $E_{DP,2D}$ and $E_{DP,3D}$ in order to conclude which version of the test produces the smallest error.

3.1.3. Outliers identification

In the last test, the user should identify, from among all the possible represented points, those highly susceptible to be considered as outliers (see fig.6).

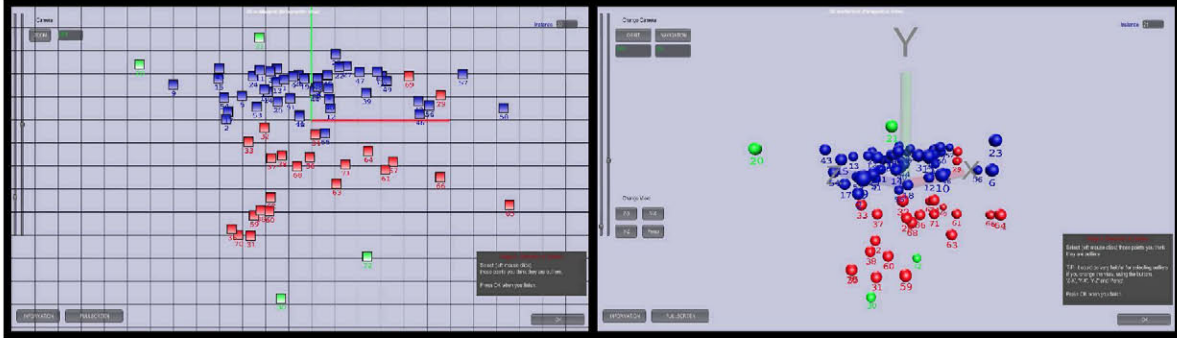


Figure 6: Outlier Identification test. The points identified as possible outliers are colored green.

Motivation. The aim is to assess the effectiveness when performing the task of outlier identification in a visual way, using a 2D and 3D scatterplot. For each user, two numerical values are obtained, $T_{OI,2D}$ and $E_{OI,2D}$. $T_{OI,2D}$ is the time taken to complete the test, and $E_{OI,2D}$ represents the % of points that the user has correctly identified as outliers (using a 3D scatterplot, the values will be $T_{OI,3D}$ and $E_{OI,3D}$). To obtain the number of points correctly identified as outliers by the user, the points truly considered outliers in the original data must be calculated previously. The *Weka* software has been used for computing the outliers.

Note: before starting the test, each user was provided with an understandable description of the definition of an *outlier*, as well as different figures illustrating various examples of outliers.

Details. As a preliminary step to the carrying out of the test, the filter '*InterquartileRange*' was used, available in *Weka*, for computing the possible outliers in the original data. A total of 13 points have been detected as potential outliers (points: 9, 10, 11, 14, 17, 20, 30, 31, 38, 39, 66, 70 and 72). Therefore, our calculations are based on these points.

Next, in the test, the user is asked about which point or points could be considered as outliers, from his point of view. As the user selected those candidate points as outliers, these points were colored green in order to distinguish them from the rest. Finally, the set of points that the user has identified as outliers are compared to those that, in fact, are outliers.

The equations for computing E_{OI2D} and E_{OI3D} are defined as:

$$E_{OI2D} = \frac{CO_{2D}}{T_O} 100; \quad (4)$$

$$E_{OI3D} = \frac{CO_{3D}}{T_O} 100; \quad (5)$$

where E_{OI2D} and E_{OI3D} are the % of points correctly identified as outliers in the 2D/3D test, respectively. Therefore, CO_{2D} and CO_{3D} represents the number of points correctly identified as outliers in the 2D/3D test, respectively. T_O is the total number of points identified as outliers in the original data and its value is 13. Note that the greater E_{OI2D} and E_{OI3D} values, the better accuracy.

3.2. Definition of the final questions

To complement the results of the visual tests, a set of questions have been also included. The users were asked these questions once they finished the tests. These questions attempt to assess the visual experience of each user with each previously performed test. The aim is to reinforce, as far as possible, the results according to the criteria and preferences of each user.

The list of questions are as follows:

- Question: From your point of view, what kind of scatterplot (2D or 3D) do you think is more useful in general to perform each of the 3 tests?. Answer: two options were available: 2 or 3. The user should select only one option. Answering this question was obligatory.
- Question: Could you tell us why?. Answer: free text answer. Answering this question was obligatory.
- Question: On a scale of 1 - 5 (5 being the best score), could you rate how comfortable you felt carrying out the test using a 2D scatterplot?. Answer: five options were available: 1, 2, 3, 4 and 5. The user should select only one option. Answering this question was obligatory.
- Question: On a scale of 1 - 5 (5 being the best score), could you rate how comfortable you felt carrying out the test using a 3D scatterplot?. Answer: five options were available: 1, 2, 3, 4 and 5. The user should select only one option. Answering this question was obligatory.
- What kind of scatterplot (2D or 3D) do you think you have been more successful in the tests, i.e., less error?. Answer: two options were available: 2 or 3. The user should select only one

option. Answering this questions was obligatory.

- Could you tell us why you think so?. Answer: free text answer. This question was not mandatory to answer.
- What kind of scatterplot (2D or 3D) did you feel more comfortable with, e.g: when you navigate through the scenes, move the camera and interact with the data points?. Answer: two options were available: 2 or 3. The user should select only one option. Answering this question was obligatory.
- Could you tell us why you think so?. Answer: free text answer. Answering this question was optional.
- Do you have any clue on how to improve the visualization in 3D or 2D scatterplots?. Answer: free text answer. Answering this question was optional.
- Finally, would you like to make any comments or suggestions about the tests for future improvement?. Answer: free text answer. Answering this question was optional.

3.3. Results

The results and a discussion about the experiments in subsections 3.1 and 3.2 are presented in the following subsections.

3.3.1. Visual tests

Table 1 presents the mean values of the results obtained during the tests, for both dimensions (2 and 3). Before analyzing these results, several aspects must be considered. Firstly, the results for each test in both dimensions will be compared. For the tests *Point Classification* (P.C) and *Outlier Identification* (O.I), the mean value of % success rate, computed on all the users, is shown (Mean $E_{PC_{2D}}$, $E_{PC_{3D}}$, $E_{OI_{2D}}$ and $E_{OI_{3D}}$). However, for the test *Distance Perception* (D.P), the total error value, computed over all the users, is shown (Total $E_{DP_{2D}}$ and $E_{DP_{3D}}$). This is because a boxplot containing the remaining information is presented below. Moreover, this total value could be considered as very significative when drawing conclusions, since it represents the total cumulative error made by all the users during that test. Lastly, the best values obtained in the tests are highlighted in bold, for the sake of clarity.

	2D			3D		
	P.C	D.P	O.I	P.C	D.P	O.I
Mean time (s)	31.797	137.757	30.605	48.332	132.732	37.398
Total time (s)	1271.11	5510.3	1224.14	1933.66	5309.28	1495.05
Mean Accuracy (%) \ Total Error	87.5	719.24	23.525	89.75	391.88	28.358

Table 1: Mean and total times values taken in the tests. Mean accuracy (in %; P.C and O.I test) and total error (D.P test) values obtained in the tests. The best values for 2 and 3 dimensions are highlighted in bold.

As regards the *Point Classification* test, it is noticed that the times taken to complete the 2D and 3D version of the test are quite different. Both the mean and the total times taken to complete the test are considerably smaller in the 2D version (Mean $T_{PC,2D}$, 31.79 s; Mean $T_{PC,3D}$, 48.33 s) than the 3D version (Total $T_{PC,2D}$ is 1271.11 s; Total $T_{PC,3D}$ is 1933.66 s). From these values it can be highlighted that the users took, on average, almost 17 seconds less when classifying points using the 2D technique, as compared to the 3D technique (see fig.7 for a better description of the time distribution). However, the mean classification rate E_{PC} remains very similar for both cases (Mean $E_{PC,2D}$, 87.5%; Mean $E_{PC,3D}$, 89.75%), obtaining a slight improvement using 3D (see fig.8B for a description of the accuracy values distribution). Therefore, the results yielded by this test suggest that the users take longer time to complete the 3D version, but this results in a little improvement in the accuracy when classifying points. It must be noted that, in this first stage, the results are not significant enough to draw any conclusions in relation to the possible improvement that could introduce the use of a third dimension to perform this task. Nevertheless, the differences in time to complete both versions of the test could outline that 3D interfaces might be further enhanced to make easier the interaction and navigation.

The *Distance Perception* test provides results radically different. Here, the users should estimate a set of proportions/rate between distances, by using the 2D and 3D scatterplot. Firstly, the total error value made when using the 2D version (Total $E_{DP,2D}$, 719.24) is 1.83 times greater than when the 3D version is used (Total $E_{DP,3D}$, 391.88). In other words, the error made by the users when perceiving and estimating the distances between the points using the 2D scatterplot is significantly greater than the error made when using the 3D version (see fig.8A). As regards the time required to complete the test, similar results are obtained in both versions. The mean time to complete the test is roughly 5 seconds less when using the 3D version (Mean $T_{DP,2D}$, 137.7 s; Mean $T_{DP,3D}$, 132.7 s). In this case, the errors made by the users when using both versions of the test yield enlightening results. This suggest that the simple fact of the inclusion of a third dimension in the data to be displayed, significantly improves the perception of the real distances between existing instances of the data, when

visualizing MMD. However, from the point of view of time, the task is performed very similarly.

The last test, *Outlier Identification*, shows results very similar to the conclusions reached in the *Point Classification* test, as both of them are based on classification concepts. When identifying different points as potential outliers using the 3D scatterplot, the users improve, on average, almost 5% in accuracy compared to the 2D version (Mean $E_{OI,2D}(\%)$, 23.52%; Mean $E_{OI,3D}(\%)$, 28.35%). However, the time taken to detect the outliers is, on average, nearly 7 seconds (Mean $T_{OI,2D}$, 30.6 s; Mean $T_{OI,3D}$, 37.3 s) less in 2D.

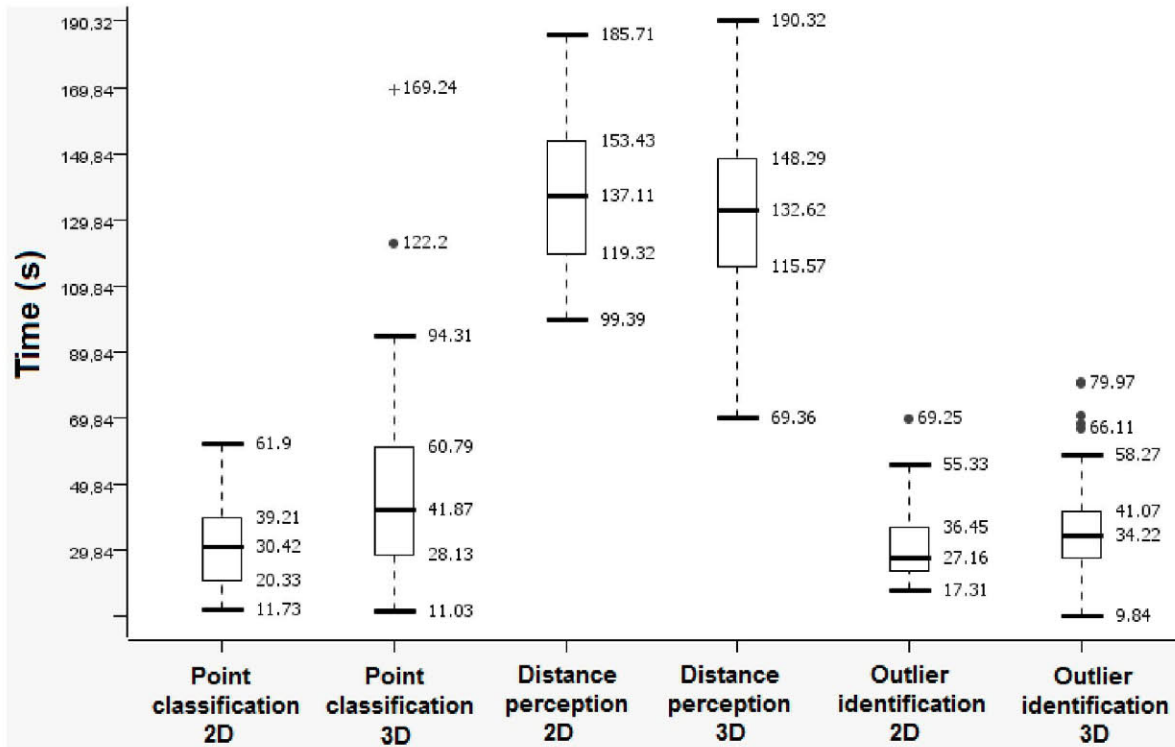


Figure 7: Distribution of the time values obtained for each of the three tests in both dimensionalities. This boxplot shows clear differences in relation to the time taken by each of the tests using the 2D and 3D scatterplot. The following points must be highlighted: less time in the realization of the 2D version of the test *Point classification*, with respect to the 3D version; the time values are slightly smaller in the realization of the 3D version of the test *Distance perception* than the 2D version; the time values are also smaller in the realization of the 2D version of the test *Outlier identification* than the 3D version.

It is worth highlighting the following fact regarding the classification results obtained in the first test. At first glance, they could be seen as high classification values (mean accuracies values of 87.5% and 89.75% for 2 and 3D, respectively), but it should be taken into account the strong influence of using a particular dataset for the study. There are datasets and studies (e.g., Leukemia data [58], by Golub et al.), in which the variables considered have been carefully selected by the authors, thereby they have a very great influence and have such discriminator power that even reducing the dimensionality

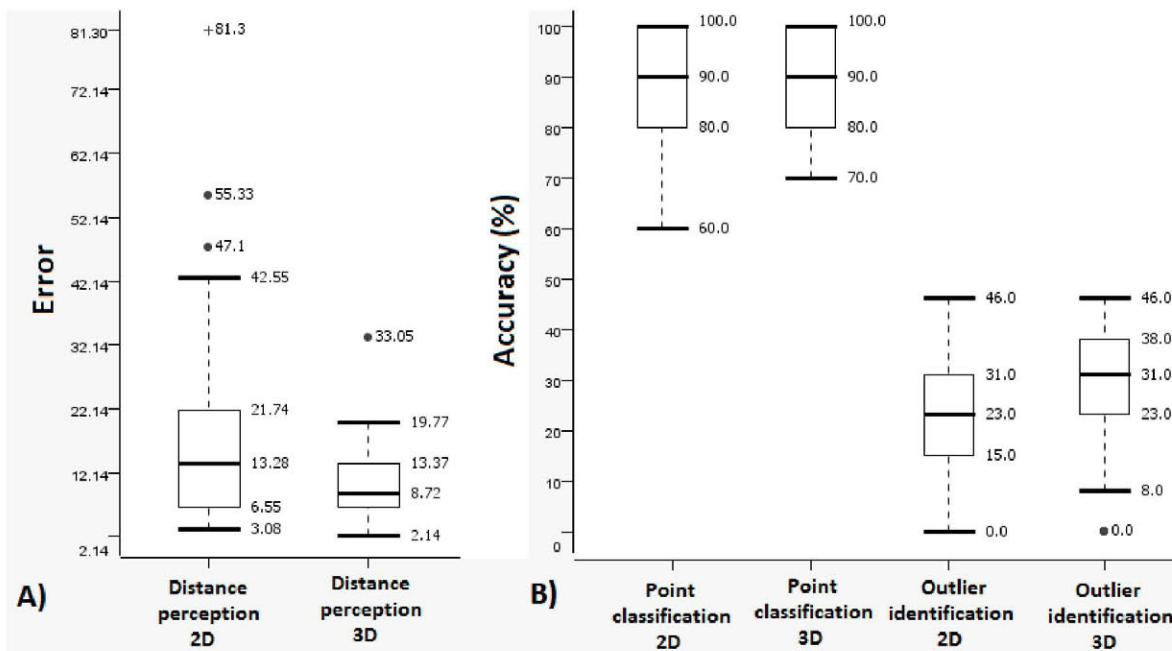


Figure 8: Distribution of the error/accuracy values obtained for each of the three tests in both dimensionalities. A: It can be clearly seen that the error values produced in the 3D version are much lower than those in the 2D version. B: The accuracy values achieved in each test do not highlighted significant differences for 2D and 3D.

of the data to very low values (as might be 2 and 3), high classification values are still obtained. Figure 9 presents an example of the accuracy values obtained when using 4 common classification algorithms (KNN [72, 73], Naive Bayes [74], Random Forest [75] and Logistic Regression [76]), when the dimensionality of the data is reduced (using PCA) along a dimensional range up to dimension 2. The results have been validated using the 0.632 *Bootstrap* method [77, 78] and exhibit significantly high values. This makes sense and confirms that common classification algorithms are able to detect, from a large number of variables, a smaller subset of variables that are significant for classification tasks, and at the same time possible redundant information is being eliminated that would hinder the classification process. Therefore, if the results obtained by users when classifying manually and those by automatic classification methods are compared, the decrease in accuracy introduced by the visual perception of the user can be highlighted.

The results presented here are not significant enough to draw definitive conclusions as regards the suitability of visualizing MMD using 3D. Nevertheless, they should be taken into account, since the improvement achieved by the inclusion of a third dimension in MMD is, in many cases, quite obvious. More specifically, the results obtained show that the changes that intrinsically occur in the accuracy (simply by using a visualization technique in two different versions, 2D and 3D) are, in some cases,

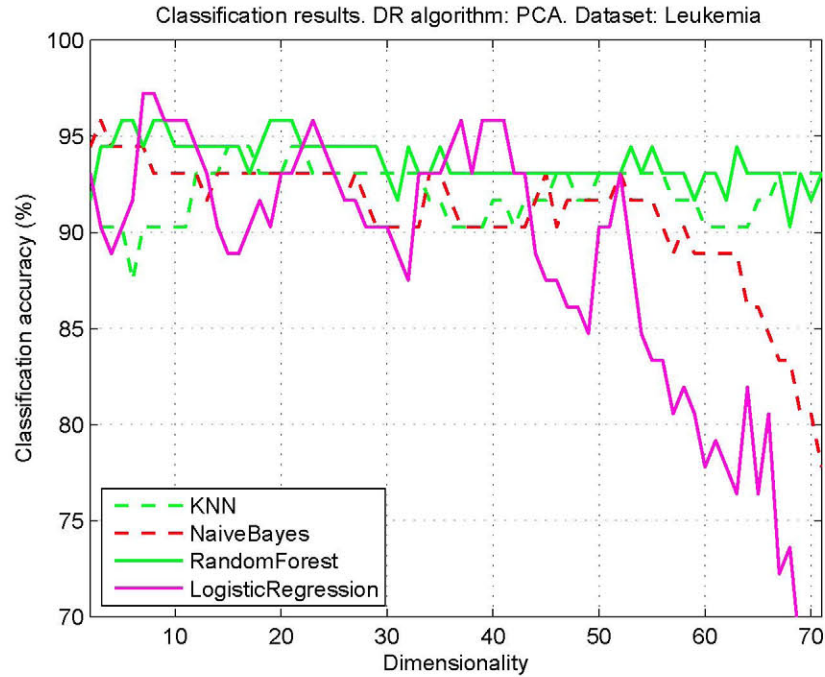


Figure 9: Accuracy values obtained by different classification methods (KNN [72, 73], Naive Bayes [74], Random Forest [75] and Logistic Regression [76]) when the dimensionality of a dataset (Leukemia) is reduced up to dimension 2, using PCA algorithm. The results have been validated using the 0.632 *Bootstrap* method [77, 78].

notorious. As regards the time needed to complete the tests, the results indicated that when the users use 2D the time is reduced, maybe because of the simplicity of 2D compared to 3D interfaces. However, at this state in the study no firm conclusions can be drawn from these results.

3.3.2. Final questions

Finally, the answers to the final questions given to each of the users who carried out the tests are shown (see subsection 3.2 and fig. 10). Firstly, in relation to *what kind of scatterplot the user thinks is more useful in general to perform each of the 3 tests*, a majority of users (55%) think that the 3D scatterplot has been more useful for carrying out the tests than the 2D version. Thus, from the users that think the *3D scatterplot* technique is more useful, some of the most repeated responses that the users answered are highlighted: when using 3D more information is available, but a good navigation through these 3 dimensions is completely necessary to be more certain of the outcome; a better appreciation of the distances between points; a greater comfort when using the different 3D views for the outlier identification test, since you easily realized that 2D points that did not seem like outliers were only so by changing the 3D views; through the similarity of 3D perception with the

human eye; because you can choose a different view plane; because it is more intuitive; it makes the spatial identification of the points easier. However, those users (45%) that think the *2D scatterplot* is more useful also highlight a preference for the 3D version of the distance test, but 2D for other tasks; in 2D there are no problems because of the perspective or occlusion data; the exploration of the place where the data are located and the establishment of distances is easier in a 2D environment; 2D interaction is simpler; in 2D there is less distortion; in 2D it is easier to perform measurements, but less accurate than if three variables are used.

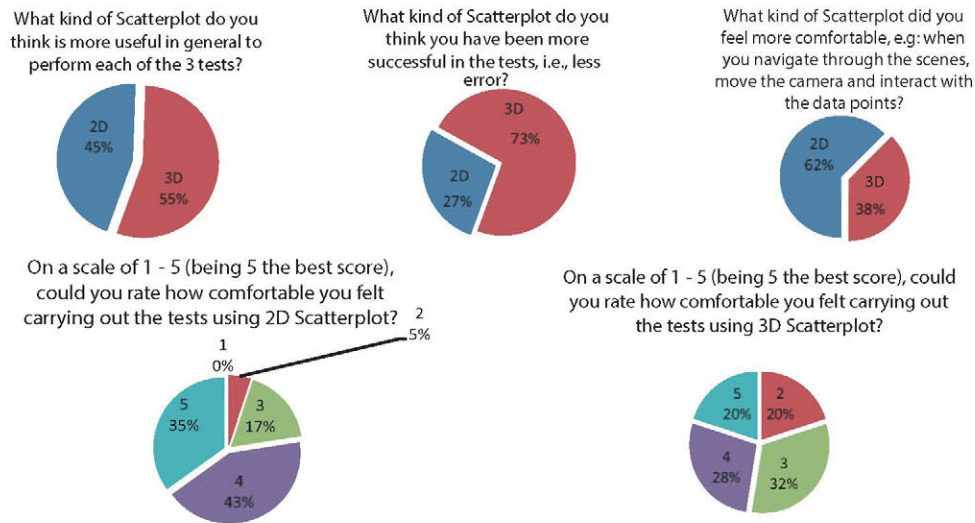


Figure 10: Users' preferences after carrying out the tests.

In relation to *what kind of scatterplot the user thinks have been more successful in the tests*, the results indicate that most of the users (73%) think that they have made an smaller error when using the *3D scatterplot*. The most repeated answers to justify this opinion are: if 3D is used, extra information is gained from the data, thus the error is smaller; the 3 dimensions help to perceive the space better; 3D is more complete in allowing data to be displayed from multiple points of view and thus it obtains a more accurate perception of them and make a smaller error; in the distance test, the 3 dimensions can correctly identify the angle between the the vectors connecting the points, thus facilitating the assessment of their relative distance. However, those users (27%) who preferred the *2D scatterplot* noted that: establishing distances in 3D is very complicated because of the perspective; the distances are easier to evaluate in 2D, since they do not depend on the position of the view; in 3D not all points can be seen at the same time; a 2D environment does not suffer from distortion because of perspective and occlusion data, unlike in 3D.

In relation to *what kind of scatterplot the user felt more comfortable with when navigating through the scenes, moving the camera and interacting with the data points*, there was a significant preference for the 2D version. The 62% of users felt more comfortable navigating through the 2D test, while the rest (38%) preferred moving through the 3D version. The justification for some of the answers that supported the 3D version are that it is more helpful in 3D since the navigation is more realistic than in 2D; it gave the impression that in 3D, the resolution was better and when zooming in and out there was a really noticeable shift in perspective, while the 2D scatterplot did not provide that feeling; a better appreciation of the real distances when navigating through the 3D scene. While the majority of the responses supporting the 2D version were: the range of movements in 3D is much more useful, but it is hard to get used to it and what offers the 2D is desirable; the 3D interaction did not work as expected. It was hard to interact and it was faster changing the view to Y-Z, X-Y and X-Z to discover the results; in 2D, the controls were simpler, easier, more comfortable and faster.

It was appropriate to complement the previous answers with several questions related to: *rate, on a scale of 1 - 5 (5 being the best score), how comfortable the user felt carrying out the test using the 2D scatterplot* and *rate, on a scale of 1 - 5, how comfortable the user felt carrying out the test using the 3D scatterplot* (Bottom figures in fig. 10). It appears that the users slightly opted for the carrying out of the tests using the 2D scatterplot, since the scores are a slightly greater. 78% of the users gave scores of 5 or 4 (35% gave scores of 5, 43% gave scores of 4) the comfort they felt when carrying out the 2D tests, whilst the 48% of the users gave scores of 5 or 4 (20% gave scores of 5, 28% gave scores of 4) when they used the 3D version.

Finally, some interesting concepts in relation to *whether the user had any clue about how to improve the visualization in 3D or 2D scatterplots* are highlighted. Some of the responses were: the inclusion of some kind of additional display, with shapes, sizes, colors and transparencies; to improve the interface for navigation in 3D; do not use red and green as colors in the same plot, since people with difficulties cannot see these colors. Maybe, there are much better color schemes available; to add a grid in order to quantify the coordinates of each point more easily; in 3D, to facilitate the operation of zooming in and out when using the perspective mode.

The results presented here refer to the second part of the users' preferences of the visual tests. On the one hand, generally most of the users think that using the 3D version of the scatterplot technique is more useful in carrying out the tasks assigned, moreover in many cases they think that the error made in the tests is smaller, a fact that actually happens.

On the other hand, the results also suggest that there is a clear and consensual trend indicating that, the users felt more comfortable carrying out the tasks when using the 2D scatterplot, mainly due to its direct and traditional use, as well as the simplicity of the 2D technique. Therefore, the conclusions outlined here highlight the fact that, there is still a lot of hard work to be done in the conception and design of appropriate, powerful and intuitive interfaces that allow the interaction in three dimensional environments when visualizing MMD using 3D visualization techniques.

There are still some clear discrepancies in the opinions of users in this second stage of the visual tests. Therefore, and similarly to the first part of the visual tests, firm conclusions still cannot be drawn that support the possible benefits of the inclusion of the third dimension to display MMD.

To summarize, the results of the *visual* tests carried out on 40 users, do not highlight definitive information on the superiority of 3D compared to 2D when visualizing MMD. However, certain advantages of using 3D visualization have already been outlined and they are worth further study. Therefore, this paper also provides an *analytical* approach that, by means of a strong mathematical background, seeks to provide new information to confirm the main hypothesis of the study. For this reason, the use of a methodology to quantify the loss of quality produced in DR tasks is proposed. Specifically for this study, the aim is to demonstrate *analytically* that the loss of quality in 2D is significantly higher than in 3D.

4. Analytical approach

Here, a quantification of the loss of quality produced in the transition from 3D to 2D on real-world datasets is carried out. This section is split into 2 subsections. The first provides a detailed definition of the methodology used to carry out the loss of quality quantification. The second subsection presents the results.

4.1. Definition

As defined in [1], the basic unity to quantify the loss of quality is:

$$Quality\ Loss = (1 - quality\ value) \tag{6}$$

where 1 represents a perfect geometry preservation, and *quality value* is the value obtained by a particular measure of quality. The domain for *quality value* is [0,1], where 0 means the worst geometry

preservation and 1 is the best possible result. The *quality loss* is the achieved *quality value* subtracted from 1. Therefore, the smaller *quality loss value*, the better geometry preservation.

Year	DR Algorithm	Reference	Package	Parameter settings
1901	Principal Component Analysis (PCA)	[62, 63]	The Matlab Toolbox for Dimensionality Reduction (2012) [79]	None (default)
1969	Sammon Mapping (SM)	[80]	The Matlab Toolbox for Dimensionality Reduction [79]	None (default)
1997	Curvilinear Component Analysis (CCA)	[81]	SOM Toolbox 2.0 (2005) [82]	$epochs = 10$ (default)
1998	Kernel PCA (KPCA)	[83, 84]	The Matlab Toolbox for Dimensionality Reduction [79]	$\kappa(x_i, x_j) = e^{-\frac{\ x_i - x_j\ ^2}{\sigma^2}}$ $\kappa(x_i, x_j) = (x_i \cdot x_j)^2$
2000	Isomap	[85, 86]	Matlab package for Isomap (MIT, 2000) [87]	$K = 7$
2000	Locally Linear Embedding (LLE)	[88, 89]	The Matlab Toolbox for Dimensionality Reduction [79]	$K = 7$
2001	Linear Discriminant Analysis (LDA)	[90, 91, 92]	The Matlab Toolbox for Dimensionality Reduction [79]	None (default)
2001	Laplacian Eigenmaps (LE)	[93, 94]	The Matlab Toolbox for Dimensionality Reduction [79]	$K = 7, \sigma = 1.0$ (default)
2004	Maximum Variance Unfolding (MVU)	[95, 96, 97]	Matlab package for MVU (2012) [98]	$K = 7$
2006	Diffusion Maps (DM)	[99, 100]	The Matlab Toolbox for Dimensionality Reduction [79]	$t = 1.0$ (default), $\sigma = 1.0$ (default)
2008	<i>t</i> -Stochastic Neighbor Embedding (<i>t</i> -SNE)	[101]	The Matlab Toolbox for Dimensionality Reduction [79]	$perplexity = 30$ (default)

Table 2: DR Algorithms used in the experiments, listed chronologically.

The methodology presented in [1] is made up of 12 DR algorithms (2 linear, 9 nonlinear) (presented in table 2) and 11 quality assessment criteria (table 3). Note that usually, by definition, for almost all the measures, 0 is the worst value and 1 is the best possible result (perfect geometry preservation). In the case of S_S , Q_M , P_M , P_{MOD} and $NIEQA_{LOCAL}$ measures it is exactly the opposite. Thus, these measures were transformed to the same semantic interpretation.

Year	Name of the measure	Reference
1969	Sammon Stress (S_S)	[80]
2001	Trustworthiness (M_T)	[102]
2001	Continuity (M_C)	[102]
2006	Local Continuity Meta-Criterion (Q_k)	[103, 104]
2007	Mean Relative Rank Errors (Q_M)	[41, 105, 106]
2009	Procrustes Measure (P_M)	[107]
2009	Modified Procrustes Measure (P_{MC})	[107]
2009	Co-ranking Matrix (Q_{NX})	[105, 108]
2009	Co-ranking Matrix (R_{NX})	[105, 108]
2011	Global Measure (Q_Y)	[109]
2012	Normalization independent embedding quality assessment ($NIEQA_{LOCAL}$)	[110]

Table 3: Selected quality assessment methods for quantifying loss of quality, listed chronologically.

Futhermore, the experiments using this methodology were performed with 12 different real-world datasets. As regards their nature, the datasets 8, 9 and 11 are exclusively of DNA microarray origin. Six of them (datasets 1, 3, 4, 5, 7 and 10) belong to other medical nature, and the rest (datasets 2, 6 and 12) to other fields. Table 4 describes the datasets.

4.1.1. Methodology

By definition, the methodology presented in [1] is based on the following steps (figure 11): dimensional thresholding computation, quality loss quantifier curves obtaining (a.k.a. QLQC, explained below), increasing/decreasing stability function and quantification analysis of loss of quality.

Dataset	Instances	Features
1. Breast Cancer Wisconsin (Diagnostic, 1995) [111]	569	30
2. Connectionist Bench (Sonar, Mines vs. Rocks, 1988) [112]	208	60
3. SPECTF Heart (2001) [113]	267	44
4. Breast Cancer Wisconsin (Prognostic, 1995) [114, 111, 115]	198	33
5. Prostate (2008) [116]	380	9
6. Glass Identification (1988) [117]	107	9
7. Parkinsons (2007) [118, 119]	195	22
8. Leukemia (1999) [58]	72	5147
9. Diffuse large B-cell lymphomas (DLBCL, 2002) [120]	77	7070
10. Gardener Classifier (neurons, 2013) [121]	241	368
11. Small Round Blue Cell Tumors (SRBCT's, 2001) [122]	83	2308
12. Libras Movement (2009) [123]	330	90

Table 4: Real-world datasets used in the experiments.

In the first step, a dimensionality interval is defined in order to quantify the loss of quality on the DR process. After that, the quality curves associated to each assessment measure are obtained. The increasing/decreasing stability function deals with the selection of those curves that meet a set of constraints. Finally, an analysis of the loss of quality over the selected curves is carried out.

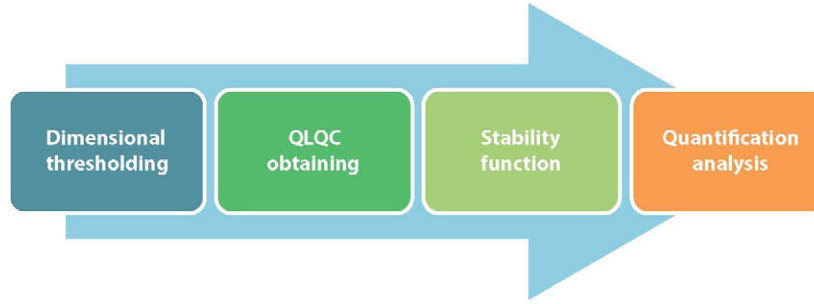


Figure 11: Methodology for quantifying the loss of quality after the DR process (taken from [1]).

Specifically, the aim of the study presented here is achieved by means the last step in the methodology, *Quantification analysis*, where a particular analysis of the loss of quality produced in specific dimensionalities can be carried out (in this case, 3 and 2 dimensions).

Dimensional thresholding computation. In order to quantify the loss of quality in the DR process, an interval is needed. An interval is achieved by two different values, so the authors defined a major and minor dimensionality threshold [1]. The *major threshold* (N') is usually limited to the original dimensionality value of the dataset to be studied. However, the minor threshold is a customizable value.

For this paper, two different values have been used for the minor threshold: 3 and 2. Therefore,

the loss of quality produced from 3D to 2D is analyzed using two intervals: $[2, N']$ and $[3, N']$.

Quality Loss Quantifier Curves (QLQC) obtaining. After defining the interval in which the loss of quality is going to be analyzed, the methodology computes a set of curves (defined by the authors as *QLQC*) as the result of evaluating the loss of quality using the 11 quality measures along the pre-defined interval (figure 12), on a particular dataset. The quality values provided by each measure can be considered as a single QLQ curve in which the X axis represents the range for dimensionalities where the data will be embedded, and the Y axis the quality value of the measurement.

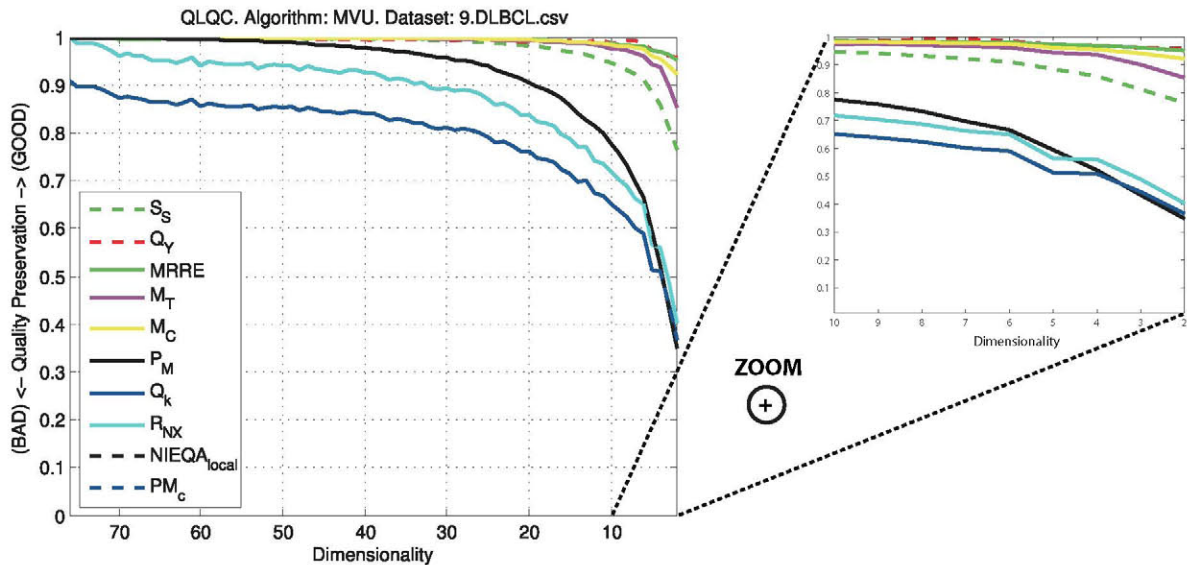


Figure 12: Example of *QLQC* plot for a particular dataset, by using a DR algorithm (MVU) (adapted with permission from [1]).

Increasing/Decreasing Stability function. As is also stated in [1], one of the main challenges was related to selecting those curves of the plot that could be useful and provide valid information when quantifying results, since drawing conclusions was not always possible. That is, they were interested in selecting those curves in which the quality values are gradual, stable and decrease (analogously, the loss of quality increases) during the DR process as the dimensionality starts from N' and progressively gets reduced until the minor threshold. The rationale for this concept is detailed in the paper.

Therefore, due to the lack of a statistical method in the literature that analyzes these concepts of *stability* or *growth* of a curve, the Increasing/Decreasing Stability function ($S_{I/D}$) was presented. $S_{I/D}$ arose in order to select those curves considered stable enough to analyze the loss of quality, and thus discard the rest of the curves that do not provide useful information (see fig. 13).

$S_{I/D} \in [-1, 1]$, where 1 represents a perfect increasing stability, 0 absence of increasing/decreasing

stability, and -1 perfect decreasing stability. It is worth mentioning that the way of discriminating between the curves considered stable and those that are not, is by using a threshold value. Thus, a threshold for $S_{I/D}$ is defined, from which it is possible to know with certainty that the selected curves meet the necessary criterion of stability.

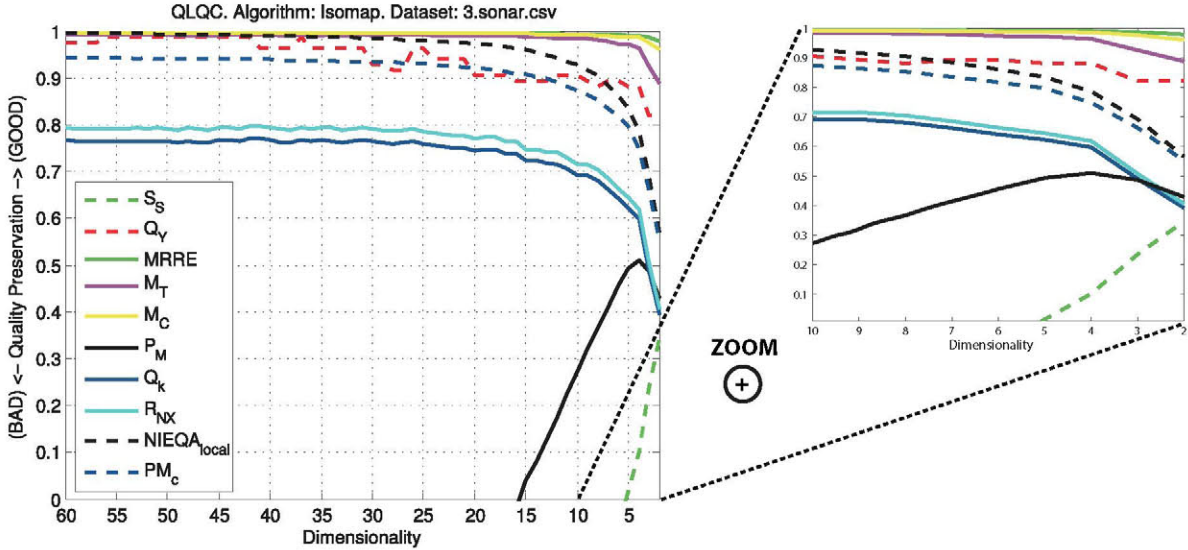


Figure 13: QLQC containing curves that violate the Increasing/Decreasing Stability criterion. The red and green dashed lines (that is, the quality curves generated by the Q_Y and S_S measures) and black line (P_M) violate the Increasing/Decreasing Stability criterion. These curves do not reach the minimum threshold to be considered suitable to analyze. The blue and light blue lines (Q_k and R_{NX} measures) present low values of Increasing/Decreasing Stability, and the rest present high values of Increasing/Decreasing Stability since they are smooth and have a decreasing behavior (reproduced with permission from [1]).

Quantification analysis of loss of quality. In the last step of the methodology, the authors proposed a quantification analysis of the loss of quality. These analyses could be from a simple analysis of the loss of quality in a certain interesting dimensionality to a more complex data analysis. Specifically, three different kinds of analysis were originally proposed [1]:

1. **Clustering of methods** according to the loss of quality throughout the entire DR process. In order to detect similar behaviors when reducing the dimensionality of the data, in terms of loss of quality, a clustering process of the DR algorithms has been carried out.
2. **Relationship** between different geometry preservation measures used in the study.
3. **Loss of quality trend** analysis from N' into B dimension. Here, the methodology represents the differences in loss of quality trend, first when the data dimensionality is reduced from N' to M , in relation to when the data dimensionality is reduced from N' to B , B being lower than M .

DR algorithms-Quality criteria	S_S	Q_M	M_T	M_C	P_M	Q_k	Q_{NX}	R_{NX}	Q_Y	$NIEQA_{LOCAL}$	P_{MC}
CCA	28.927	X	3.662	X	X	2.922	3.138	3.138	1.553	13.612	8.332
DM	30.9818	23.488	30.644	38.356	33.457	10.695	11.803	11.803	X	26.687	17.117
ISOMAP	X	9.289	16.482	12.556	X	3.622	3.957	4.167	X	13.805	8.098
$KPCA_{gouss}$	1.913	6.821	9.153	10.294	4.200	3.432	3.791	3.791	X	9.892	6.038
$KPCA_{poly}$	X	7.218	16.065	8.010	X	2.456	3.451	3.451	2.062	26.726	10.287
LAPLACIAN	X	4.036	4.738	5.924	X	1.876	2.059	2.059	-1.298	6.597	3.067
LDA	1.871	1.998	3.592	2.223	X	X	3.328	X	3.360	8.744	4.030
LLE	17.846	10.887	12.740	8.704	X	X	5.162	X	X	10.067	4.177
MVU	47.660	29.418	28.910	33.326	38.388	12.625	16.470	18.277	X	45.195	39.688
PCA	28.145	34.261	27.871	30.073	40.364	9.567	12.953	12.953	14.285	35.890	48.617
SM	46.775	41.070	44.212	42.864	36.002	18.536	24.189	24.189	X	40.649	37.971
t -SNE	X	23.420	10.806	16.668	X	4.480	5.4764	5.476	X	19.828	30.560

Table 5: Mean values (in %, each value is the mean of the $Q.L.R_{3D \rightarrow 2D}$ values obtained on each of the 12 datasets) of loss of quality in the transition from 3D to 2D (e.g., S_S obtains a value of 28% when reducing the dimensionality using CCA. This means that the S_S measure quantifies a mean loss of quality value of 28.92% only in the transition from 3D to 2D regarding the total loss of quality from $N'D$ to 2D). X values represent when there are no computed values on any of the datasets, due to technical restrictions on the algorithms used in the methodology.

As said above, this paper focuses on a particular case of the last kind of analysis (number 3), *Loss of quality trend*, where the B and M values have been set to 2 and 3, respectively. In this way, it is possible to quantify the loss of quality produced in the transition from 3D to 2D.

4.2. Results

This subsection is completely focused on the results obtained from the experiments (completely performed in Matlab) in terms of the loss of quality produced in the transition from 3D to 2D.

Table 5 and figure 14 show the mean values of loss of quality reported by each quality criterion, when reducing from 3D to 2D using a particular DR algorithm. The values are in % and they represent, of the total amount of loss of quality produced from $N'D$ to 2D, which is the mean percentage of loss of quality generated only in the transition from 3D to 2D. This mean value is the mean loss of quality computed on all the datasets. The higher the values, the stronger the loss of quality reported between 3D and 2D spaces.

Particularly, the method for calculating each value in the table is summarized as follows: the mean loss of qualities from $N'D$ to 2D are computed, that is, loss of quality value in $N'D$, in $(N' - 1)D$, and so on up to 2D. After that, the mean of these values is obtained and called *total loss of quality*. It is an indicator of how is the transition in the loss of quality throughout the whole DR process. The second step is exactly the same as the previous one but, instead of 2D, the loss of qualities up to 3D are computed (it is called *3D quality loss*). The final value (in %) is the ratio between both values:

$$Quality\ LossRatio(Q.L.R)_{3D \rightarrow 2D} = \left(1.0 - \frac{3D\ quality\ loss}{total\ quality\ loss}\right) * 100; \quad (7)$$

Rewriting formally what was said above, each value in Table 5 represents the mean of the $Q.L.R_{3D \rightarrow 2D}$

values on all the datasets, when reducing the data using a particular DR method and measuring the loss of quality through a quality criterion.

As regards the X values in Table 5, it means that it has not been possible to obtain results on any of the datasets, due to the technical issues of the DR algorithms and quality criteria used in the methodology. However, it is considered that the rest of the results presented here involve enough experimentation on several datasets to provide firm results in the quantification process.

It is worth mentioning that before analyzing the results of the quantification, Table 5 provides interesting information about the stability (in terms of technical restrictions of the algorithm used) of the quality criteria, as well as the DR algorithms. If the table is observed at a column level the best quality criteria, in terms of stability on all the datasets, are M_T , Q_{NX} , $NIEQ_{LOCAL}$ and P_{MC} (No X values in columns, thus they always obtained results). Q_M and M_C criteria also present good values for stability, since they rarely failed when producing results. However, P_M , S_S and Q_Y measures were quite unstable in all the datasets, as they often failed. If the table at a row level is analyzed, the best DR algorithms are PCA, DM, MVU, SM and $KPCA_{gauss}$. However, the worst results are obtained by Isomap, LLE, t -SNE, LDA and CCA.

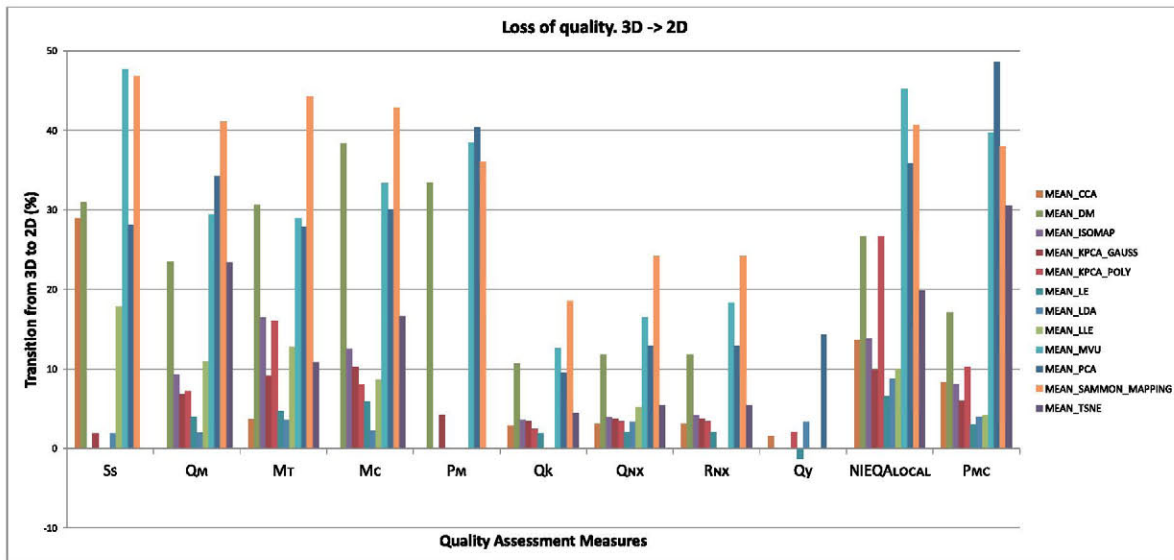


Figure 14: Mean loss of quality values in the transition from 3D to 2D (results from Table 5). X axis represents how the different quality criteria quantify the loss of quality, when reducing the data dimensionality from 3D to 2D using the different DR algorithms on all the datasets. Y axis shows the mean loss of quality values. The data are presented in a scale 0%-50%.

The following subsections analyze the loss of quality from two different approaches: the first one is a point of view from the quality criteria, and the other is from the DR algorithms. In both cases,

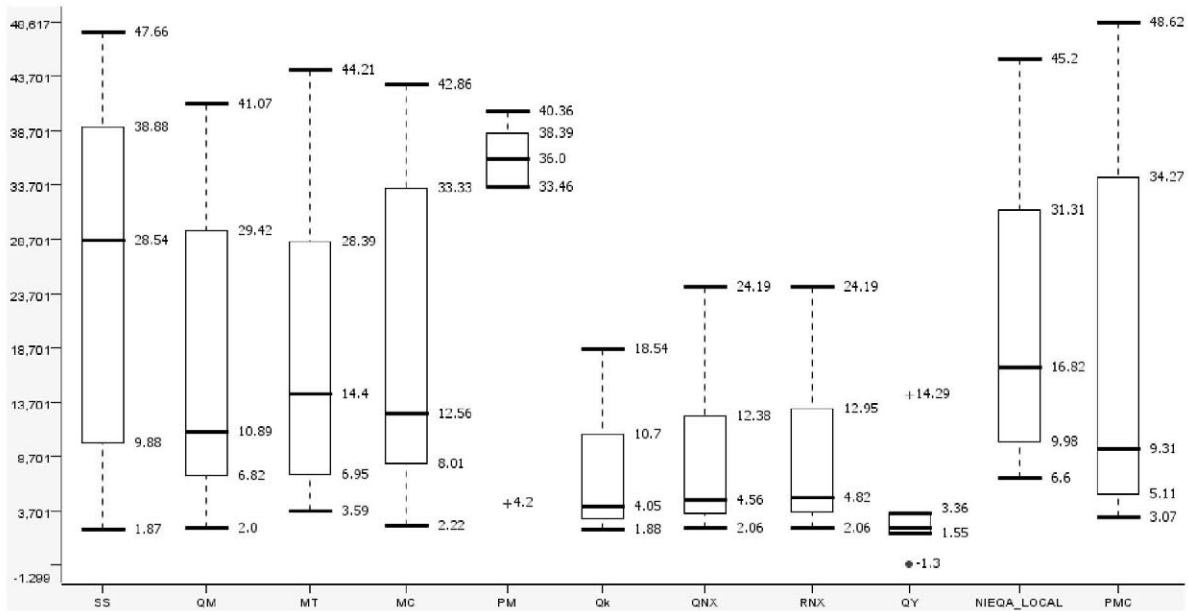


Figure 15: Boxplot that shows the distribution of the mean loss of quality values at quality criteria level (boxplots correspond to columns in Table 5). The data are presented in a scale 0%-50%. This represents to what extent each quality criterion quantifies the loss of quality, for all the DR algorithms.

boxplots are used for making easier the interpretation of the results.

4.2.1. Quality criteria

Firstly, according to figure 15, it is worth noting the disparity between distributions of some quality criteria. According to the different distribution of the quality values reported by each criterion, several groups could be observed: 1) Q_Y ; 2) Q_k , Q_{NX} and R_{NX} ; 3) Q_M , M_T , M_C , $NIEQA_{LOCAL}$ and P_{MC} ; 4) S_S ; 5) P_M .

Firstly, the Q_Y criterion reports low and a very different distribution of loss of qualities as regards the rest of the criteria (around a median value of 1.55% and 3.36% of maximum value). This is due to its unique way of evaluating the loss of quality, since it is not based on comparable concepts to the rest of the measures and it involves both local and global concepts. Furthermore, the boxplot indicates outliers (14.29% and -1.3% values) for the Q_Y criterion, that provide further information about its previously cited instability.

The quality criteria in the second group produce median values of around 5%, minimum values of 2%, and maximum values of around 24% of loss of quality. This fact could be explained by the different nature of conception of each of those criteria, as they are based on a mechanism for assessing the loss of quality by locality concepts based on the ranking of nearest neighbors.

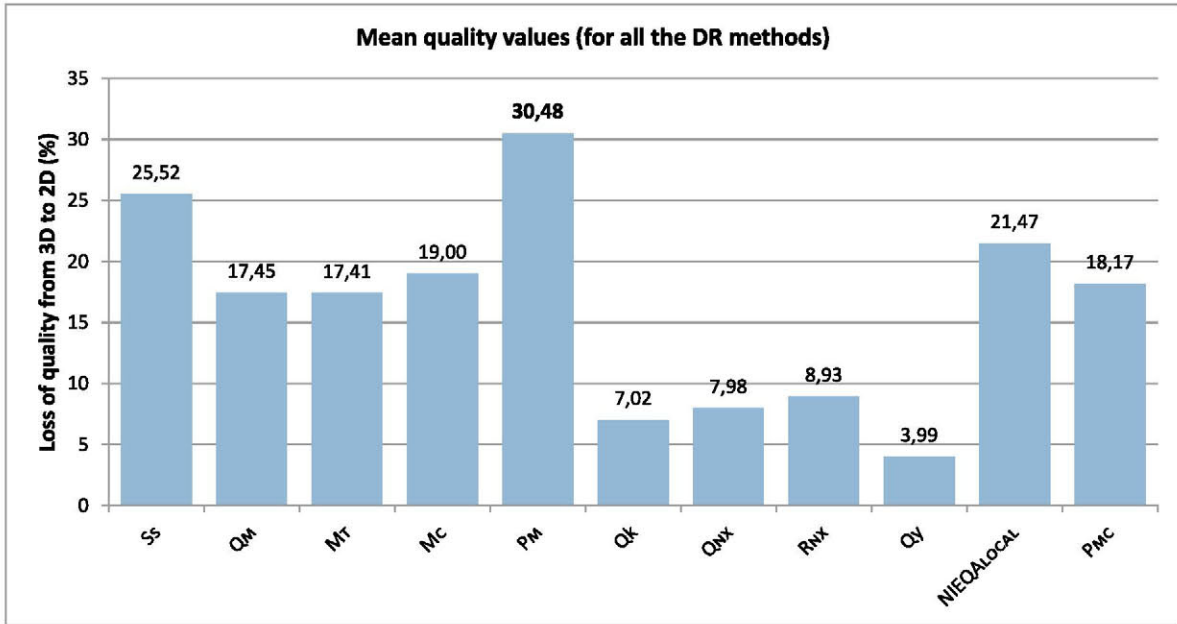


Figure 16: Mean quality values reported by the quality criteria for all the DR algorithms. It is quite clear that, for almost all the quality criteria, the mean values of loss of quality in the transition from 3D to 2D are high enough to be taken into account. The data are presented in a scale 0%-35%. The highest loss of quality value is highlighted in bold.

The third group, group criteria that quantifies a high loss of quality in the transition from 3D to 2D, varying from 9.31% in P_{MC} , up to 16.82% in $NIEQA_{LOCAL}$ for median values and reaching maximum values of more than 41.07% for all these criteria, which are, indeed, very high values. In turn, it is worth mentioning that this group is divided into two subgroups according to the different nature of the quality criteria contained in it: Q_M , M_T and M_C measures are based on neighborhood overlapping concepts and show a similar behaviour when capturing the loss of quality; however, $NIEQA_{LOCAL}$ and P_{MC} measures use procrustes analysis techniques as background mechanism.

The S_S criterion reported a median value of 28.54%, minimum value of 1.87% and reached a maximum value of 47.66%, which represent high values indeed. This measure, unlike the rest, is based on global concepts to quantify the loss of quality.

The last group is represented by the P_M criterion, as the distribution of its quality values is quite different from the rest. It reported a minimum value of 33.46%, a median value of 36% and a maximum value of 40.36% of loss of quality. As can be seen, these losses of quality are very significant. It is worth highlighting that P_M , P_{MC} and $NIEQA_{LOCAL}$ have very similar concepts of development, since they use procrustes analysis techniques. However, P_M and P_{MC} behave slightly differently, as P_M was originally meant to assess data generated by isometric DR algorithms (such as Isomap) and P_{MC} also

works with normalized DR algorithms (e.g. PCA). Taking into account this constraint for P_M when assessing normalized embeddings, it might be that some values are being modified or skewed in a disproportionately way. This fact is corroborated when analyzing the stability of P_M on the datasets, since its unstability when working with normalized embeddings is high. Furthermore, the boxplot for P_M reports outliers (4.2% value).

4.2.2. DR algorithms

When analyzing the loss of quality at DR algorithm level (Figure 17), the following fact can be observed: the DR algorithms that generate a greater loss of quality from 3D to 2D are SM, MVU, PCA, DM and t -SNE. This could be explained by the fact that when reducing the dimensionality from N D to 3D, the loss of quality is not very significant, however in the transition from 3D to 2D a substantial increase in the loss of quality occurs with respect to the loss of quality produced up to 3D. Particularly, of the total amount of loss of quality generated when reducing from N D to 2D, a great percentage (median values of 39.31%, 31.37%, 28.15%, 25.09% and 13.74% respectively for SM, MVU, PCA, DM and t -SNE; maximum values of 48.62%, 47.66%, 46.78%, 38.36% and 30.56% respectively for PCA, MVU, SM, DM and t -SNE) only occurs in the switch from 3D to 2D. This could suggest that when using these algorithms for DR purposes, the first three features (or dimensions) have so far recovered the majority of the original information contained within the initial dataset and,

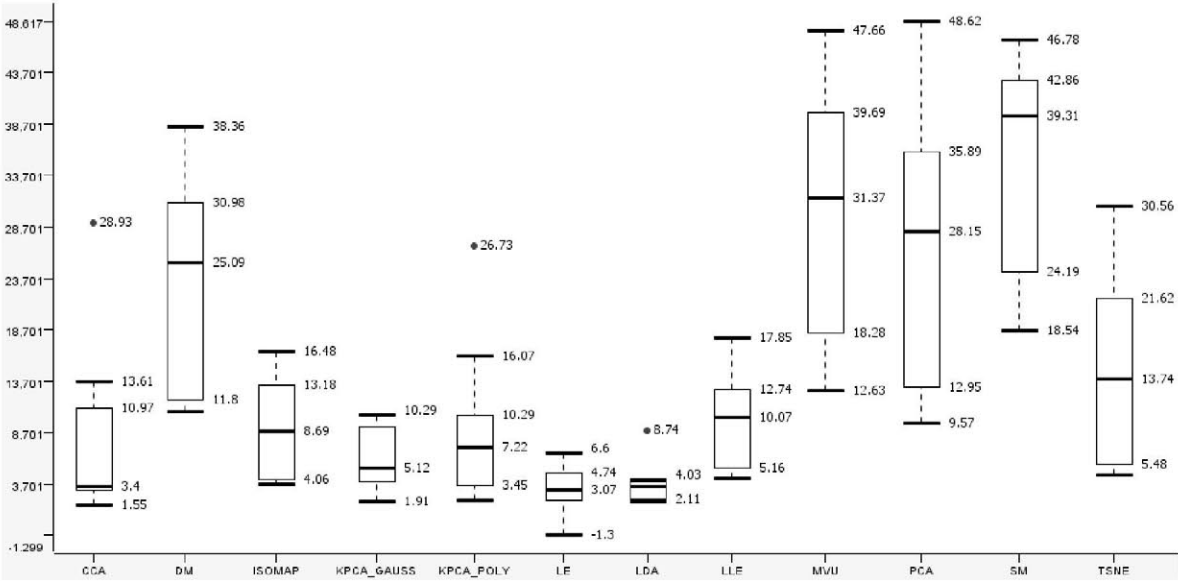


Figure 17: Boxplot that shows the distribution of the mean loss of quality values at DR algorithm level (each boxplot corresponds to a row in Table 5). The data are presented in a scale 0%-50%.

from there on, the recovery of information is considerably slower. It is also noticed that, for all these algorithms, the variance in the distribution of the reported values is significantly higher than the rest of the algorithms.

LDA, LE, CCA and $KPCA_{gauss}$ algorithms score low median values of loss of quality in the transition from 3D to 2D, as well as distributions with low variance. In addition, according to Figure 17, CCA and LDA algorithms show distributions with outliers. This coincides with the unstable behaviours for these aforementioned algorithms. However, the maximum values of loss of quality achieved by two of these algorithms, CCA and $KPCA_{gauss}$, are high (13.61% and 10.29%, respectively).

The rest of the algorithms (Isomap, $KPCA_{poly}$ and LLE) also report high values. Specifically, median values from 7.22% to 10.07%, and maximum values around 17%.

To sum up, both approaches (subsections 4.2.1 and 4.2.2) show that the loss of quality in 2D spaces far exceed that which occurs in 3D spaces. To be more precise, the theoretical results indicate that the loss of quality, when switching from 3D to 2D, reaches maximum values of 48.62% (see figures 15 and 17) and mean values of 30.483% (see figure 16) of the total loss of quality for many cases, which can be considered noticeably high.

5. Conclusions and discussions

How to visualize data is an important question, especially for MMD with more than two attributes this seems to be an open issue in VA, human computer interaction, and computer graphics in general. The simplicity and intuition provided by DV techniques in 2D spaces is certainly one key to their success. However, the aim of this paper is to demonstrate scientifically that the use of three dimensions on the visualization counteracts the benefits of dealing with traditional 2D. From a point of view based on the loss of quality the results are conclusive, 3D showed a solid and significant superiority over 2D visualization. In this sense, few times before this concept had been analyzed and quantified in this particular way.

To prove the superiority of 3D over 2D when visualizing MMD, first, a battery of tests on a sample of 40 users attempts to demonstrate statistically, by means of visualization, whether conclusions on the improvement in the accuracy and efficiency produced by the inclusion of a third dimension can be drawn. Secondly, an analytical quantification of the loss of quality produced when reducing the dimensionality of the data from 3D to 2D is proposed in order to yield new insights into the possible superiority of the third dimension. This quantification is done by using a recently proposed

methodology.

The tests in the visual statistical approach showed that the error made by the users carrying out a set of tasks in DV is considerably smaller when visualizing MMD in 3D, specially in distances assessment. This could suggest a greater accuracy when working with the data using 3D visualization. As regards the time taken to complete each of the tests proposed, it is observed that the users needed less time to complete the 2D version of the tests. So, at this point no firm conclusions about the superiority of 3D visualization could be drawn. As regards the users' preferences and suggestions the results indicated that, and taking into account the clear improvement and work still needed in the development of 3D displays, working with 3 dimensions may be equally or even more helpful than traditionally done using 2D DV. But there were some clear discrepancies in the opinions of users and thus strong conclusions still cannot be drawn about the benefits of 3D to display MMD.

Nevertheless, the results obtained through the analytical approach showed that the average and maximum loss of qualities obtained only when reducing the data dimensionality from 3 to 2, are 30.483% and 48.62% (respectively) of the total loss of quality produced throughout the whole DR process (from the original dimensionality of the data to 2D). This means that a high degree of loss of quality occurs just passing from 3D to 2D, which can make us reconsider whether that 2D reduction is really necessary or not.

These results provide definitive conclusions, as well as a demonstration of the superiority of using a 3D environment when MMD are visualized. The concept of quality degradation could be crucial when visualizing data, and it is demonstrated that the loss of quality produced with the inclusion of a third dimension is noticeably smaller than just using 2 dimensions. This fact strongly suggests the suitability of the third dimension for embedding and visualizing MMD, as well as for manifold learning tasks where the intrinsic dimensionality of the dataset is unknown or greater than 2. Therefore, this fact allow the original hypothesis to be confirmed and it should be taken into account for future developments.

6. Future lines

There is still a lot of work to be done in the improvement of 3D displays and interfaces for addressing the drawbacks associated with the use of a third dimension. Therefore, in the near future, we are interested in developing a 3D interface for reducing the aforementioned drawbacks. To do so, a set of usability tests by using different user profiles (e.g. experts in different fields and non-computer people)

could be carried out. This interface should also allow 2D MMDV, by advising users the loss of quality involved in this transition. We could also apply this interface to classified data (in a supervised and unsupervised way). The idea is to perform a classification process of new data, based on previously classified data.

Finally, the research presented here opens up a wide range of possibilities for performing a more profound comparative study of the DR algorithms according their geometry preservation skills, as well as the most important quality assessment criteria. Furthermore, the results of this initial study could be extended to analyse whether there is a significant improvement in the classification accuracy, by using 3 instead of 2 dimensions for supervised learning.

Acknowledgements

The authors thankfully acknowledge the computer resources, technical expertise and assistance provided by the Centro de Supercomputación y Visualización de Madrid (CeSViMa), the Spanish Supercomputing Network and the Graphics-Interactive Systems Group at TU Darmstadt, Germany. The study was supported by the Spanish Ministry of Economy and Competitiveness (Grant TIN2010-21289-C02-02 and the Cajal Blue Brain Project, Spanish partner of the Blue Brain Project initiative from EPFL).

References

- [1] Antonio Gracia, Santiago González, Víctor Robles, and Ernestina Menasalvas. A methodology to compare dimensionality reduction algorithms in terms of loss of quality. *Information Sciences*, 270(0):1 – 27, 2014.
- [2] Patrick E. Hoffman and Georges G. Grinstein. Information visualization in data mining and knowledge discovery. chapter A Survey of Visualizations for High-dimensional Data Mining, pages 47–82. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002.
- [3] Pak Chung Wong and R. Daniel Bergeron. 30 Years of Multidimensional Multivariate Visualization, 1997. Chapter 1 (pp. 3–33) of Gregory M. Nielson, Hans Hagen, and Heinrich Müller, editors, *Scientific Visualization: Overviews, Methodologies, and Techniques*, IEEE Computer Society.

- [4] Melanie Tory and Torsten Möller. A Model-Based Visualization Taxonomy, 2002. Technical Report CMPT-TR2002-06, Computing Science Department, Simon Fraser University.
- [5] Daniel A. Keim. Designing Pixel-Oriented Visualization Techniques: Theory and Applications. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):59–78, January 2000.
- [6] Daniel A. Keim. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, January 2002.
- [7] Daniel A. Keim and H.-P. Kriegel. Visualization Techniques for Mining Large Databases: A Comparison. *Transactions on Knowledge and Data Engineering, Special Issue on Data Mining*, 8(6):923938, 1996.
- [8] J.A. Hartigan. Printer graphics for clustering. *Journal of Statistical Computation and Simulation*, 4(3):187–213, 1975.
- [9] George W. Furnas and Andreas Buja. Prosection Views: Dimensional Inference through Sections and Projections. *Journal of Computational and Graphical Statistics*, 3(4):323–385, 1994.
- [10] L. Tweedie and R. Spence. The Prosection Matrix: A Tool to Support the Interactive Exploration of Statistical Models and Data. *Computational Statistics*, 13:65–76, 1998.
- [11] Jarke J. van Wijk and Robert van Liere. HyperSlice - Visualization of Scalar Functions of Many Variables. In Gregory M. Nielson and R. Daniel Bergeron, editors, *IEEE Visualization*, pages 119–125. IEEE Computer Society, 1993.
- [12] Pak Chung Wong, Andrew H. Crabb, and R. Daniel Bergeron. Dual multiresolution HyperSlice for multivariate data visualization. In Stuart K. Card, Stephen G. Eick, and Nahum D. Gershon, editors, *INFOVIS*, pages 74–75. IEEE, 1996.
- [13] B. Alpern and L. Carter. The Hyperbox. In *Proc. of Visualization'91*, pages 133–139, San Diego, CA, 1991.
- [14] A. Inselberg. Multidimensional Detective. In *Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97)*, INFOVIS '97, pages 100–, Washington, DC, USA, 1997. IEEE Computer Society.
- [15] Alfred Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.

- [16] D. F. Andrews. Plots of High-Dimensional Data. *Biometrics*, 28(1), 1972.
- [17] P. E. Hoffman. *Table Visualizations: A Formal Model and Its Applications*. PhD thesis, Computer Science Department, University of Massachusetts at Lowell, 1999.
- [18] J. Beddow. Shape Coding of Multidimensional Data on a Microcomputer Display. In *IEEE Visualization*, pages 238–246, 1990.
- [19] Ramana Rao and Stuart K. Card. The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus + Context Visualization for Tabular Information. pages 318–322. ACM, 1994.
- [20] G. Peano. Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen*, 36:157–160, 1890.
- [21] David Hilbert. Ueber die stetige Abbildung einer Line auf ein Flächenstück. *Mathematische Annalen*, 38(3):459–460, 1891.
- [22] Morton. A computer oriented geodetic data base and a new technique in file sequencing. Technical Report Ottawa, Ontario, Canada, 1966.
- [23] Daniel A. Keim, Mihael Ankerst, and Hans-Peter Kriegel. Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data. In *Proceedings of the 6th Conference on Visualization '95, VIS '95*, pages 279–, Washington, DC, USA, 1995. IEEE Computer Society.
- [24] Daniel A. Keim and Hans-Peter Kriegel. VisDB: Database Exploration Using Multidimensional Visualization. *IEEE Comput. Graph. Appl.*, 14(5):40–49, 1994.
- [25] Mihael Ankerst, Daniel A. Keim, and Hans-Peter Kriegel. Circle Segments: A Technique for Visually Exploring Large Multidimensional Data Sets. In *Visualization '96, Hot Topic Session, 1996*, 1996.
- [26] Daniel A. Keim, Ming Hao, Umesh Dayal, Meichun Hsu, and Julain Ladisch. Pixel Bar Charts: A New Technique for Visualizing Large Multi-Attribute Data Sets Without Aggregation. In *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, INFOVIS '01, pages 113–, Washington, DC, USA, 2001. IEEE Computer Society.

- [27] T. Mihalisin, E. Gawlinski, J. Timlin, and J. Schwegler. Visualizing a Scalar Field on an N-dimensional Lattice. In *Proceedings of the 1st Conference on Visualization '90*, VIS '90, pages 255–262, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.
- [28] T. Mihalisin, J. Timlin, and J. Schwegler. Visualization and Analysis of Multi-variate Data: A Technique for All Fields. In *Proceedings of the 2Nd Conference on Visualization '91*, VIS '91, pages 171–178, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.
- [29] Ted Mihalisin, John Timlin, and John Schwegler. Readings in Information Visualization. chapter Visualizing Multivariate Functions, Data, and Distributions, pages 115–125. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.
- [30] Jeffrey LeBlanc, Matthew O. Ward, and Norman Wittels. Exploring N-dimensional Databases. In *Proceedings of the 1st Conference on Visualization '90*, VIS '90, pages 230–237, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.
- [31] S. K. Feiner and Clifford Beshers. Visualizing N-dimensional Virtual Worlds with N-vision. *SIGGRAPH Comput. Graph.*, 24(2):37–38, 1990.
- [32] Ben Shneiderman. Tree Visualization with Tree-maps: 2-d Space-filling Approach. *ACM Trans. Graph.*, 11(1):92–99, 1992.
- [33] Maria Cristina Ferreira de Oliveira and Haim Levkowitz. From Visual Data Exploration to Visual Data Mining: A Survey. *IEEE Trans. Vis. Comput. Graph.*, 9(3):378–394, 2003.
- [34] Herman Chernoff. The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973.
- [35] J.M. Chambers, W.S. Cleveland, B. Kleiner, and P.A. Tukey. Graphical Methods for Data Analysis. *The Wadsworth Statistics/Probability Series. Boston, MA: Duxury*, 1983.
- [36] Ronald M. Pickett and Georges G. Grinstein. Iconographic Displays For Visualizing Multidimensional Data. *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics*, 1:514–519, 1988.
- [37] Haim Levkowitz. Color Icons: Merging Color and Texture Perception for Integrated Visualization of Multiple Parameters. In *IEEE Visualization*, pages 164–170, 1991.

- [38] Alexei A. Efros and William T. Freeman. Image Quilting for Texture Synthesis and Transfer. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 341–346, New York, NY, USA, 2001. ACM.
- [39] Li-Yi Wei and Marc Levoy. Fast Texture Synthesis Using Tree-structured Vector Quantization. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 479–488, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
- [40] Colin Ware and William Knight. Using Visual Texture for Information Display. *ACM Trans. Graph.*, 14(1):3–20, 1995.
- [41] John A. Lee and Michel. Verleysen. *Nonlinear dimensionality reduction*. Springer, New York; London, 2007.
- [42] Daniel A. Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Information visualization. chapter Visual Analytics: Definition, Process, and Challenges, pages 154–175. Springer-Verlag, Berlin, Heidelberg, 2008.
- [43] J. Wang. *Geometric Structure of High-dimensional Data and Dimensionality Reduction*. Higher Education Press, 2012.
- [44] Mark F. St. John, Michael B. Cowen, Harvey S. Smallman, and Heather M. Oonk. The Use of 2D and 3D Displays for Shape-Understanding versus Relative-Position Tasks. *Human Factors*, 43(1):79–98, 2001.
- [45] C. Ware. *Visual Thinking: for Design*. Morgan Kaufmann series in interactive technologies. Elsevier Science, 2010.
- [46] Tovi Grossman, Daniel Wigdor, and Ravin Balakrishnan. Exploring and reducing the effects of orientation on text readability in volumetric displays. In Mary Beth Rosson and David J. Gilmore, editors, *CHI*, pages 483–492. ACM, 2007.
- [47] K.F Van Orden and J.W Broyles. Visuospatial Task Performance as a Function of Two and Three-Dimensional Display Presentation Techniques. *Displays*, 21(1):17–24, 2000.
- [48] Sung Ha Park and Jeffrey C. Woldstad. Multiple Two-Dimensional Displays as an Alternative to Three-Dimensional Displays in Telerobotic Tasks. *Human Factors*, 42(4):592–603, 2000.

- [49] Melanie Tory, Arthur E. Kirkpatrick, M. Stella Atkins, and Torsten Möller. Visualization Task Performance with 2D, 3D, and Combination Displays. *IEEE Transactions on Visualization and Computer Graphics*, 12(1):2–13, 2006.
- [50] Melanie Tory, David W. Sprague, Fuqu Wu, Wing Yan So, and Tamara Munzner. Spatialization Design: Comparing Points and Landscapes. *IEEE Trans. Vis. Comput. Graph.*, 13(6):1262–1269, 2007.
- [51] Melanie Tory, Colin Swindells, and Rebecca Dreezer. Comparing Dot and Landscape Spatializations for Visual Memory Differences. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1033–1040, 2009.
- [52] Harvey S. Smallman, Mark F. St. John, Heather M. Oonk, and Michael B. Cowen. Information Availability in 2D and 3D Displays. *IEEE Computer Graphics and Applications*, 21(5):51–57, 2001.
- [53] S.I. Fabrikant. *Spatial Metaphors for Browsing Large Data Archives*. University of Colorado, 2000.
- [54] C.D Wickens. The proximity compatibility principle: Its psychological foundation and relevance to display design. Technical report, ARL-92-5/NASA-92-3. Savoy, IL: University of Illinois Institute of Aviation, Aviation Research Lab, 1992.
- [55] C.D Wickens. Virtual reality and education. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 1992*, pages 842–847, 1992.
- [56] C.D Wickens, D.H Merwin, and E.L Lin. Implications of graphics enhancements for the visualization of scientific data: dimensional integrality, stereopsis, motion, and mesh. *Human Factors*, 36(1):44–61, 1994.
- [57] Michael Sedlmair, Tamara Munzner, and Melanie Tory. Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2634–2643, 2013.
- [58] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, 1999.

- [59] Sandrine Dudoit, Jane Fridlyand, and Terence P. Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457):77–87, 2002.
- [60] Chen-Hsiang Yeang, Sridhar Ramaswamy, Pablo Tamayo, Sayan Mukherjee, Ryan M. Rifkin, Michael Angelo, Michael Reich, Eric S. Lander, Jill P. Mesirov, and Todd R. Golub. Molecular classification of multiple tumor types. In *ISMB (Supplement of Bioinformatics)*, 2001, pages 316–322, 2001.
- [61] Sayan Mukherjee, Pablo Tamayo, Simon Rogers, Ryan M. Rifkin, Anna Engle, Colin Campbell, Todd R. Golub, and Jill P. Mesirov. Estimating Dataset Size Requirements for Classifying DNA Microarray Data. *Journal of Computational Biology*, 10(2):119–142, 2003.
- [62] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.
- [63] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, 24, 1933.
- [64] W. Goldstone. *Unity 3.x Game Development Essentials*. Community experience distilled. Packt Publishing, 2011.
- [65] Carlos-Miguel Lorenzo. A review of interoperability and possibilities for data analysis from virtual world environments. *Int. J. Metadata Semant. Ontologies*, 6(3/4):234–240, July 2011.
- [66] MJ Kim, X Wang, PED Love, H Li, and SC Kang. Virtual reality for the built environment: a critical review of recent advances. *ITcon*, 18:279–305, 2013.
- [67] A. M. Manferdini and F Remondino. A review of reality-based 3d model generation, segmentation and web-based visualization methods. *International Journal of Heritage in the Digital Era*, 1(1):103–124, 2012.
- [68] Carolin Helbig, Hans-Stefan Bauer, Karsten Rink, Volker Wulfmeyer, Michael Frank, and Olaf Kolditz. Concept and workflow for 3d visualization of atmospheric data in a virtual reality environment for analytical approaches. *Environmental Earth Sciences*, pages 1–14, 2014.
- [69] Karsten Rink, Lars Bilke, and Olaf Kolditz. Visualisation strategies for environmental modelling data. *Environmental Earth Sciences*, pages 1–12, 2014.

- [70] A. Cioc, S.G. Djorgovski, C. Donalek, E. Lawler, F. Sauer, and G. Longo. Data visualization using immersive virtual reality tools. In *American Astronomical Society Meeting Abstracts*, volume 221 of *American Astronomical Society Meeting Abstracts*, January 2013.
- [71] Carina Ellinor Irene Westling, Leon Barker, Aoife Healy, Robert Needham, Harry Witchel, and Nachiappan Chockalingam. Spatial data correlation: an interactive 3d visualisation tool for correlating the motion capture data streams from difference devices. In *Electronic Visualisation and the Arts (EVA 2013)*, London, 2013. British Computing Society.
- [72] T. Cover and P. Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.
- [73] T. Mitchell. *Machine Learning (Mcgraw-Hill International Edit)*. McGraw-Hill Education (ISE Editions), 1st edition, October 1997.
- [74] Marvin Minsky. Steps toward artificial intelligence. In Edward A. Feigenbaum and Jerome A. Feldman, editors, *Computers and Thought*, pages 406–450. McGraw-Hill, New York, 1963.
- [75] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [76] D.W. Hosmer, S. Lemeshow, and R.X. Sturdivant. *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley, 2013.
- [77] B. Efron. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *Journal of the American Statistical Association*, 78(382):316–331, 1983.
- [78] Bradley Efron and Robert Tibshirani. Improvements on cross-validation: The .632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.
- [79] Laurens van der Maaten. *The Matlab Toolbox for Dimensionality Reduction*, 2012.
- [80] J. W. Sammon. A Nonlinear Mapping for Data Structure Analysis. *IEEE Transactions on Computers*, C-18(5), 1969.
- [81] P. Demartines and J. Herault. Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Trans. Neural Netw.*, 8(1):148–154, 1997.
- [82] Juha Vesanto, Johan Himberg, Esa Alhoniemi, and Juha Parhankangas. *SOM Toolbox 2.0*, 2005.

- [83] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.*, 10(5):1299–1319, July 1998.
- [84] Bernhard Schölkopf, Alexander J. Smola, and Klaus R. Müller. *Kernel principal component analysis*, pages 327–352. MIT Press, Cambridge, MA, USA, 1999.
- [85] J.B. Tenenbaum, V. Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [86] Vin De Silva and Joshua B. Tenenbaum. Global Versus Local Methods in Nonlinear Dimensionality Reduction. In *Advances in Neural Information Processing Systems 15, 2003*, volume 15, pages 705–712, 2003.
- [87] J.B Tenenbaum. Matlab package for isomap. stanford. (2000, accessed 10 december 2013).
- [88] Sam T. Roweis and Lawrence K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.
- [89] Lawrence K. Saul and Sam T. Roweis. Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. *Journal of Machine Learning Research*, 4:119–155, June 2003.
- [90] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification*. Pattern Classification and Scene Analysis: Pattern Classification. Wiley, 2001.
- [91] T.J. Hastie, R.J. Tibshirani, and J.J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics Series. Darinka Springer & Janez, 2001.
- [92] Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [93] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In *Advances in Neural Information Processing Systems, 2001*, volume 14, pages 585–591, 2001.
- [94] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396, June 2003.

- [95] K.Q. Weinberger, F. Sha, and L. K. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In *Proceedings of the Twenty First International Conference on Machine Learning (ICML-04), 2004*, pages 839–846, Banff, Canada, 2004.
- [96] Kilian Q. Weinberger and Lawrence K. Saul. Unsupervised Learning of Image Manifolds by Semidefinite Programming. *Int. J. Comput. Vision*, 70(1):77–90, October 2006.
- [97] Killan Q. Weinberger and Lawrence K. Saul. An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In *proceedings of the 21st national conference on Artificial intelligence - Volume 2, AAAI 2006*, pages 1683–1686, 2006.
- [98] Kilian Q. Weinberger. Matlab package for mvu. (2012, accessed 12 december 2013).
- [99] Boaz Nadler, Stéphane Lafon, Ronald R. Coifman, and Ioannis G. Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113–127, 2006.
- [100] Stéphane Lafon and Ann B. Lee. Diffusion Maps and Coarse-Graining: A Unified Framework for Dimensionality Reduction, Graph Partitioning, and Data Set Parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1393–1403, 2006.
- [101] Laurens van der Maaten and Geoffrey Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- [102] Jarkko Venna and Samuel Kaski. Local multidimensional scaling. *Neural Networks*, 19(6-7):889–899, 2006.
- [103] Lisha Chen. *Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Layout and Proximity Analysis, Ph.D. Thesis*. University of Pennsylvania, 2006.
- [104] Lisha Chen and Andreas Buja. Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Drawing, and Proximity Analysis. *Journal of the American Statistical Association*, 104(485):209–219, 2009.
- [105] John A. Lee and Michel Verleysen. Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomput.*, 72(7-9):1431–1443, 2009.
- [106] John Aldo Lee and Michel Verleysen. Rank-based quality assessment of nonlinear dimensionality reduction. In *ESANN, 2008*, pages 49–54, 2008.

- [107] Yair Goldberg and Ya'Acov Ritov. Local procrustes for manifold embedding: a measure of embedding quality and embedding algorithms. *Machine Learning*, 77(1):1–25, oct 2009.
- [108] John Aldo Lee and Michel Verleysen. Quality assessment of nonlinear dimensionality reduction based on K-ary neighborhoods. *Journal of Machine Learning Research - Proceedings Track*, 4:21–35, 2008.
- [109] Deyu Meng, Yee Leung, and Zongben Xu. A new quality assessment criterion for nonlinear dimensionality reduction. *Neurocomputing*, 74(6):941–948, 2011.
- [110] Peng Zhang, Yuanyuan Ren, and Bo Zhang. A new embedding quality assessment method for manifold learning. *Neurocomputing*, 97:251–266, 2012.
- [111] Olvi L. Mangasarian, W. Nick Street, and William H. Wolberg. Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43:570–577, 1995.
- [112] Paul R. Gorman and Terrence J. Sejnowski. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*, 1(1):75–89, 1988.
- [113] Lukasz A Kurgan, Krzysztof J Cios, Ryszard Tadeusiewicz, Marek R Ogiela, and Lucy S Goodenday. Knowledge discovery approach to automated cardiac SPECT diagnosis. *Artificial Intelligence in Medicine*, 23:149, 2001.
- [114] W. Nick Street, Olvi L. Mangasarian, and W. H. Wolberg. An inductive learning approach to prognostic prediction. In *ICML, 1995*, pages 522–530, 1995.
- [115] William H. Wolberg, W. Nick Street, DM Heisey, and Olvi L. Mangasarian. Computerized breast cancer diagnosis and prognosis from fine-needle aspirates. *Archives of Surgery*, 130(5):511–516, 1995.
- [116] Sunita R. Setlur, Kirsten D. Mertz, Yujin Hoshida, Francesca Demichelis, Mathieu Lupien, Sven Perner, Andrea Sboner, Yudi Pawitan, Ove Andrén, Laura A. Johnson, Jeff Tang, Hans-Olov Adami, Stefano Calza, Arul M. Chinnaiyan, Daniel Rhodes, Scott Tomlins, Katja Fall, Lorelei A. Mucci, Philip W. Kantoff, Meir J. Stampfer, Swen-Olof Andersson, Eberhard Varenhorst, Jan-Erik Johansson, Myles Brown, Todd R. Golub, and Mark A. Rubin. Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. *Journal of the National Cancer Institute*, 100(11):815–825, June 2008.

- [117] I. W. Evett and E. J. Spiehler. Knowledge Based Systems. chapter Rule induction in forensic science, pages 152–160. Halsted Press, New York, NY, USA, 1988.
- [118] Max A. Little, Patrick E. Mcsharry, Stephen J. Roberts, Declan A. E. Costello, and Irene M. Moroz. Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. *BioMedical Engineering OnLine*, 6(1):23+, 2007.
- [119] Max A. Little, Patrick E. McSharry, Eric J. Hunter, Jennifer L. Spielman, and Lorraine O. Ramig. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson’s Disease. *IEEE Trans. Biomed. Engineering*, 56(4):1015–1022, 2009.
- [120] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. S. Pinkus, T. S. Ray, M. A. Koval, K. W. Last, A. Norton, T. A. Lister, J. Mesirov, D. S. Neuberg, E. S. Lander, J. C. Aster, and T. R. Golub. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med*, 8(1):68–74, January 2002.
- [121] J. DeFelipe, P.L. Lopez-Cruz, R. Benavides-Piccione, C. Bielza, P. Larrañaga, and et al. New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nature Reviews Neuroscience*, 14(3):202–216, 2013.
- [122] Javed Khan, Jun S. Wei, Markus Ringner, Lao H. Saal, Marc Ladanyi, Frank Westermann, Frank Berthold, Manfred Schwab, Cristina R. Antonescu, Carsten Peterson, and Paul S. Meltzer. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7(6):673–679, 2001.
- [123] Daniel B. Dias, Renata C. B. Madeo, Thiago Rocha, Helton H. Bísvaro, and Sarajane M. Peres. Hand movement recognition for brazilian sign language: a study using distance-based neural networks. In *Proceedings of the 2009 international joint conference on Neural Networks, IJCNN’09*, pages 2355–2362. IEEE Press, 2009.