Visual Data Mining of Astronomic Data with Virtual Reality Spaces: Understanding the Underlying Structure of Large Data Sets

Valdés, Julio

National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

Visual Data Mining of Astronomic Data with Virtual Reality Spaces: Understanding the Underlying Structure of Large Data Sets *

Valdés, J.
October 2004

* published at the Astronomical Data Analysis Software & Systems XIVADASS XIV, Conference Proceedings. October 24 - 27, 2004. Pasadena, California, USA. NRC 47391.

Canada

Astronomical Data Analysis Software and Systems XIV **O1-2**
ASP Conference Series, Vol. XXX, 2005
P. L. Shopbell, M. C. Britton, and R. Ebert, eds.

Visual Data Mining of Astronomic Data with Virtual Reality Spaces: Understanding the Underlying Structure of Large Data Sets

Julio J. Valdés

National Research Council Canada
Institute for Information Technology
1200 Montreal Rd.
Ottawa, ON K1A 0R6
julio.valdes@nrc.ca

Abstract. The information explosion in astronomy requires the development of data mining procedures that speed up the process of scientific discovery, and the in-depth understanding of the internal structure of the data. This is crucial for the identification of valid, novel, potentially useful, and understandable patterns (regularities, oddities, etc).
A Virtual Reality (VR) approach for large heterogeneous, incomplete and imprecise information is introduced for the problem of visualizing and analyzing astronomic data. The method is based on mappings between one heterogeneous space representing the data, and a homogeneous virtual reality space. This VR-based visual data mining technique allows the incorporation of the unmatched geometric capabilities of the human brain into the knowledge discovery process, and helps in understanding data structure and patterns. This approach has been applied successfully to a wide variety of real-world domains, and it has a large potential in astronomy. Examples are presented from the domain of galaxy research.

1. Introduction

The science of astronomy has experienced unprecedented progress in the last years. In particular, the advances in computer, communication, and observation technologies have increased in many orders of magnitude the quantity and quality of astronomic data. This information explosion requires the development of data mining procedures that speed up the process of scientific discovery, and the in-depth understanding of the internal structure of the data. This is crucial for the non-trivial process of identifying valid, novel, potentially useful, and ultimately *understandable patterns* in data; that is, for knowledge discovery (Fayyad et.al 1996).

The information explosion requires analytic and interpretation procedures which enable users to *understand* their data rapidly and with greater ease. Further, the increasing complexity of the data analysis procedures makes it more difficult for the user to extract useful information out of the results given by the various techniques applied. Visual techniques are, therefore, very appealing.

In general, objects under study are described in terms of collections of *heterogeneous* properties. For example, an astronomic source can be characterized by a set of properties represented by nominal, ordinal or real-valued variables (scalar), as well as by other of a more complex nature like images (in the visible wavelength region, infrared, and others), time-series (e.g. spectra), etc. In addition, the information comes with different degrees of precision, uncertainty and completion (missing data is quite common). Classical data mining and analysis methods are sometimes difficult to use, the output of many procedures may be large and time consuming to analyze, and often their interpretation requires special expertise. Moreover, some methods are based on assumptions about the data which limit their application, specially for the purpose of exploration, comparison, hypothesis formation, etc, typical of the first stages of scientific investigation.

This makes graphical representation directly appealing. Humans perceive most of the information through vision, in large quantities and at very high input rates. The human brain is extremely well qualified for the fast understanding of complex visual patterns, and still outperforms the computer. Several reasons make Virtual Reality (VR) a suitable paradigm: Virtual Reality is *flexible*, as it allows the choice of different representation models to better suit different human perception preferences. It allows the construction of different virtual worlds representing *the same* underlying information, but with different look and feel. Thus, the user can choose the most appealing representation. VR allows *immersion*. The user can navigate inside the data, and interact with the objects in the world. VR creates a *living* experience. The user is not merely a passive observer or an outsider, but an actor in the world, in fact, part of the information itself. VR is *broad and deep*. The user may see the VR world as a whole, and/or concentrate the focus of attention on specific details or portions of the world. Of no less importance is the fact that in order to interact with a Virtual World only minimal skills are required.

In this paper a *Virtual Reality* approach for understanding large heterogeneous, incomplete and imprecise data (Valdés 2002, 2002b, Valdés & Bonham-Carter 2003, Valdés 2003, Valdés 2004), is introduced in the domain of astronomy. In this approach, the notion of data is not restricted to databases, but also includes logical relations and other forms of structured knowledge.

2. The Heterogeneous Space

Consider an *information system* $S =< U, A >$ where U and A are non-empty finite sets, called the *universe* and the set of *attributes* respectively, such that each $a \in A$ has a domain V_a and an evaluation function f_a assigns to each $u \in U$ an element $f_a(u) \in V_a$ (i.e. $f_a(u) : U \rightarrow V_a$) (here the V_a are not required to be finite). An example is shown in Fig 1. There are attributes with domains of different kinds (nominal, ordinal, ratio, fuzzy, images, time-series and graphs), and also containing missing values (represented as ?).

Heterogeneous and incomplete information systems will be considered as follows. Let `?` be a special symbol having two basic properties: *i*) if $? \in \Omega$ (Ω being an arbitrary set) and f is any unary function defined on Ω, then $f(?) = ?$, and *ii*) `?` is an incomparable element w.r.t any ordering relation defined on Ω. A

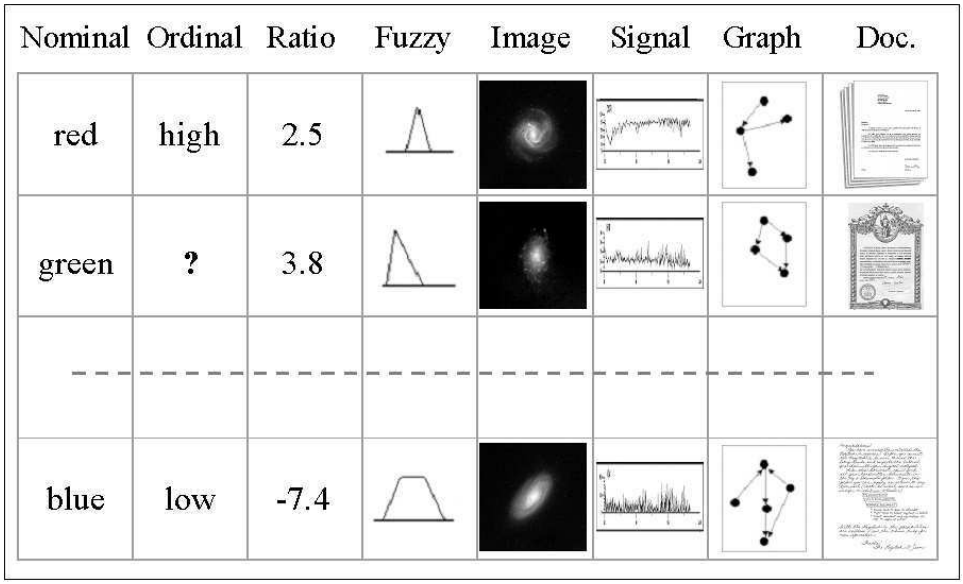

Figure 1. An example of a heterogeneous database.

heterogeneous domain is defined as a Cartesian product of a collection of *source sets* (Ψ_i): $\hat{\mathcal{H}}^n = \Psi_1 \times \ldots \times \Psi_n$, where $n > 0$ is the number of *information sources* to consider.

As an example, consider the case of a heterogeneous domain where objects are characterized by attributes given by continuous crisp quantities, discrete features, fuzzy features, graphs and digital images. Let $\mathbf{R}$ be the reals with the usual ordering, and $\mathcal{R} \subseteq \mathbf{R}$. Now define $\hat{\mathcal{R}} = \mathcal{R} \cup \{?\}$ to be a source set and extend the ordering relation to a partial order accordingly ($\hat{\mathcal{R}}$ may model scalar measurements, with missing values). Now let $\mathbf{N}$ be the set of natural numbers and consider a family of n_r sets ($n_r \in \mathbf{N}^+ = \mathbf{N} - \{0\}$) given by $\hat{\mathcal{R}}^{n_r} = \hat{\mathcal{R}}_1 \times \ldots \times \hat{\mathcal{R}}_{n_r}$ (n_r times) where each $\hat{\mathcal{R}}_j$ ($0 \leq j \leq n_r$) is constructed as $\hat{\mathcal{R}}$, and define $\hat{\mathcal{R}}^0 = \phi$ (the empty set). Now let $\mathcal{O}_j$, $1 \leq j \leq n_o \in \mathbf{N}^+$ be a family of finite sets with cardinalities k_j^o respectively, composed by arbitrary elements, such that each set has a fully ordering relation $\leq_{\mathcal{O}_j}$. Construct the sets $\hat{\mathcal{O}}_j = \mathcal{O}_j \cup \{?\}$, and for each of them define a partial ordering $\hat{\leq}_{\mathcal{O}_j}$ by extending $\leq_{\mathcal{O}_j}$ according to the definition of ?. Analogously construct the set $\hat{\mathcal{O}}^{n_o} = \hat{\mathcal{O}}_1 \times \ldots \times \hat{\mathcal{O}}_{n_o}$ (n_o times and $\hat{\mathcal{O}}^0 = \phi$). For the special case of nominal variables, let $\mathcal{N}_j$, $1 \leq j \leq n_m$ ($n_m \in \mathbf{N}^+$) be a family of finite sets with cardinalities $k_j^m \in \mathbf{N}^+$ composed by arbitrary elements but such that no ordering relation is defined on any of the $\mathcal{N}_j$ sets. Now construct the sets $\hat{\mathcal{N}}_j = \mathcal{N}_j \cup \{?\}$, and define $\hat{\mathcal{N}}^{n_m} = \hat{\mathcal{N}}_1 \times \ldots \times \hat{\mathcal{N}}_{n_m}$, ($n_m$ times and $\hat{\mathcal{N}}^0 = \phi$). Sets $\hat{\mathcal{O}}^{n_o}$, $\hat{\mathcal{N}}^{n_m}$ may represent the case of n_o ordinal variables and n_m nominal variables respectively. Similarly, a collection of n_f extended fuzzy sets $\hat{\mathcal{F}}_j$ ($1 \leq j \leq n_f$) , n_g extended graphs $\hat{\mathcal{G}}_j$ ($1 \leq j \leq n_g$) and n_i extended digital images $\hat{\mathcal{I}}_j$ ($1 \leq j \leq n_i$), can be used for constructing the corresponding cartesian products given by $\hat{\mathcal{F}}^{n_f}$, $\hat{\mathcal{G}}^{n_g}$

and $\hat{\mathcal{I}}^{n_i}$.
The heterogeneous domain is given by $\hat{\mathcal{H}}^n = \hat{\mathcal{R}}^{n_r} \times \hat{\mathcal{O}}^{n_o} \times \hat{\mathcal{N}}^{n_m} \times \hat{\mathcal{F}}^{n_f} \times \hat{\mathcal{G}}^{n_g} \times \hat{\mathcal{I}}^{n_i}$. Elements of this domain will be objects $o \in \hat{\mathcal{H}}^n$ given by tuples of length $n = n_r + n_o + n_m + n_f + n_g + n_i$, with $n > 0$ (the empty set is excluded). Other kinds of heterogeneous domains can be constructed in a similar manner, using the appropriate source sets. More general information systems are those in which the universe is endowed with a set of relations of different arities. Let $t =< t_1, \ldots, t_p >$ be a sequence of p natural integers, called *type*, and $\underline{Y} =< Y, \gamma_1, \ldots, \gamma_p >$ a relational structure as defined in (Hajek & Havranek 1978), where Y is a non-empty domain of objects and the $\Gamma = \{\gamma_i\}$ $(i = 1, \ldots, p)$ are different relations of various arities defined on Y (according to t). The extended information system will be $\hat{S} =< U, A, \Gamma >$, endowed with the relational system $\underline{U} =< U, \Gamma >$.

3. The Virtual Reality Space

A *virtual reality space* is a structure composed of different sets and functions defined as $\Upsilon =< \underline{O}, G, B, \Re^m, g_o, l, g_r, b, r >$. $\underline{O}$ is a relational structure defined as above ($\underline{O} =< O, \Gamma^v >$, $\Gamma^v =< \gamma_1^v, \ldots, \gamma_q^v >$, $q \in \mathbf{N}^+$ and the $o \in O$ are objects), G is a non-empty set of *geometries* representing the different objects and relations (the *empty* or *invisible* geometry is a possible one). B is a non-empty set of *behaviors* (i.e. ways in which the objects from the virtual world will express themselves: movement, response to stimulus, etc.). $\Re^m \subset \mathbf{R}^m$ is a *metric space* of dimension m (euclidean or not) which will be the actual virtual reality geometric space. The other elements are mappings: $g_o : O \to G$, $l : O \to \Re^m$, $g_r : \Gamma^v \to G$, $b : O \to B$, r is a collection of characteristic functions for Γ^v, $(r_1, \ldots, r_q)$ s.t. $r_i : \gamma_i^{v\,t_i} \to \{0, 1\}$, according to the type t associated with Γ^v.

The representation of an extended information system $\hat{S}$ in a virtual world implies the construction of another $\hat{S}^v =< O, A^v, \Gamma^v >, \underline{O}$ in Υ, which requires the specification of several sets and a collection of extra mappings (w.r.t. those required for Υ). A desideratum for $\hat{S}^v$ is to keep as many properties from $\hat{S}$ as possible. Thus, a natural requirement is that U and O are in one-to-one correspondence (with a mapping $\xi : U \to O$). The structural link is given by a mapping $f : \hat{\mathcal{H}}^n \to \Re^m$. If $u =< f_{a_1}(u), \ldots, f_{a_n}(u) >$ and $\xi(u) = o$, then $l(o) = f(\xi(< f_{a_1}(u), \ldots, f_{a_n}(u) >)) =< f_{a_1^v}(o), \ldots, f_{a_m^v}(o) >$ ($f_{a_i^v}$ are the evaluation functions of A^v). This gives *semantics* to the pair $< g_o(o), l(o) >$ (it determines important properties like geometry, visibility and location).
It is natural to require that $\Gamma^v \subseteq \Gamma$ (possibly empty), thus having a virtual world portraying selected relations from the information system, represented according to the choices made for G and g_r.

4. The Problem of Large Datasets

Regardless of the criteria followed when computing a virtual reality space, complex optimization procedures are applied involving the estimation of the image of the data objects. The objective function surface becomes more complex and

convoluted with the increase of the dimensionality of the parameter space, and local extrema entrapment is typical. Even if all of the difficulties related with the amount of memory and the numeric computation involved are put aside (note that a dissimilarity matrix grows quadratically with the number of objects), the graphical representation of millions or possibly billions of objects in a screen with the current computer technologies, is neither feasible, nor practical. On the other hand, assuming that it would be possible, the amount of information presented to the user will be overwhelming, and will obscure, rather than clarify, the presence of meaningful or interesting patterns. The approach followed here is to study the properties of the dataset ($\mathbf{X}$), possibly huge, in order to extract a subset of a sufficiently smaller cardinality which will either retain as much structural information as possible, or guarantee its preservation up to a predefined threshold. In this approach only the non-redundant objects up to a predefined degree are preserved, thus producing a kernel or core representation of the original dataset. If a similarity measure S is chosen as a redundancy criterium, and a similarity threshold T_s is set forth as a parameter, it is possible to construct a set $\mathbf{L} \subseteq \mathbf{X}$, such that $\forall x \in \mathbf{X}, \exists\ l \in \mathbf{L}, S(x,l) \geq T_s$ (Fig-2). There are efficient algorithms which can generate $\mathbf{L}$-sets at different T_s-levels, and this parameter will determine both the cardinality of the resulting $\mathbf{L}$-set, as well as its semantics. According to this approach, a VR representation of a large or huge dataset is obtained by first extracting a $\mathbf{L}$-set according to a suitable similarity threshold, and then computing its VR space. Since each of the data objects is represented by a sufficiently similar l-object (lower bounded by T_s), the VR space is compliant with the similarity structure of the whole dataset $\mathbf{X}$ at that level.

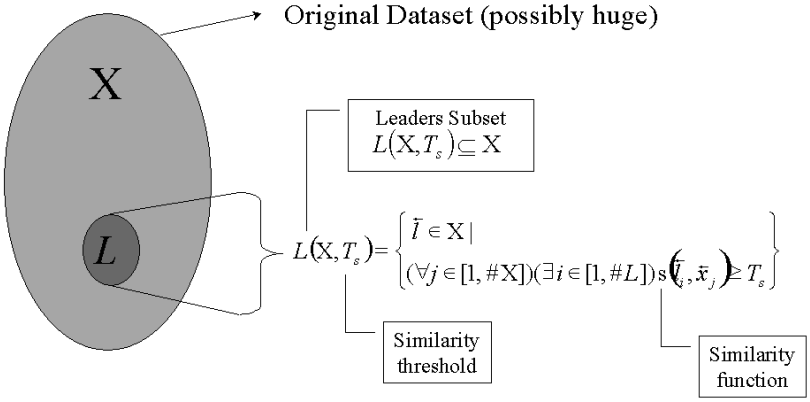

Figure 2. Relation between a dataset $\mathbf{X}$ and its corresponding $\mathbf{L}$-subset at the T_s-similarity level (# denotes set cardinality).

4.1. The Direct and Inverse Transforms

As mentioned, f plays an important role in giving semantics to the virtual world, and there are many ways in which such a mapping can be defined. In a great extent it depends on which features from the original data need to be highlighted. In particular, adjacency relationships between the objects O in Υ

should give an indication about the *similarity relationships* (Chandon, Pinson 1981) between the objects in the original heterogeneous space $\hat{\mathcal{H}}^n$ (Valdés 2002b). Other interpretations about internal structure are related with the linear/non-linear separability of class membership relations defined on the data (Jianchang & Jain 1995). In this sense, f can be constructed according to several criteria: *i)* to maximize some metric/non-metric structure preservation criteria as has been done for decades in multidimensional scaling (Kruskal 1964), (Borg & Lingoes 1987), *ii)* minimize some error measure of information loss, *iii)* maximize some measure of class separability (in a supervised case), or *iv)* satisfy several criteria simultaneously. For example, in the case of *i)*, if δ_{ij} is a dissimilarity measure between any two $i, j \in U$ ($i, j \in [1, N]$, where n is the number of objects), and $\zeta_{i^v j^v}$ is another dissimilarity measure defined on objects $i^v, j^v \in O$ from Υ ($i^v = \xi(i), j^v = \xi(j)$, they are in one-to-one correspondence), two examples of error measures frequently used are:

$$S\ stress = \sqrt{\frac{\sum_{i<j} (\delta_{ij}^2 - \zeta_{ij}^2)^2}{\sum_{i<j} \delta_{ij}^4}} \quad (1)$$

$$Sammon\ error = \frac{1}{\sum_{i<j} \delta_{ij}} \frac{\sum_{i<j} (\delta_{ij} - \zeta_{ij})^2}{\delta_{ij}} \quad (2)$$

The f mappings obtained using approaches of this kind are only implicit, as no functional representations are found, and its usefulness depends the final errors obtained in the optimization process. Explicit mappings can be obtained from these solutions using neural network, genetic programming, and other techniques. An explicit f is useful for both practical and theoretical reasons. On one hand, in dynamic data sets (e.g. systems being monitored or databases formed incrementally from continuous processes) an explicit direct transform f will speed up the incremental update of the VR information system S^v. On the other hand, it can give semantics to its attributes, thus acting as a dimensionality reducer or as a generator of new attributes.

The possibilities derived from this approach are practically unlimited, since the number of different similarity, dissimilarity and distance functions definable is immense. Moreover, similarities and distances can be transformed into dissimilarities according to a wide variety of schemes. This provides a rich framework where one can find appropriate measures better suited to both the internal structure of the data, and external criteria.

The existence of an *inverse transformation* f^{-1} from Υ back to $\hat{\mathcal{H}}^n$ is, in many cases, worth considering. If a sense is made of patterns of objects in Υ in terms of abstract concepts, and new conjectured objects or relations are conceived, it is natural to ask what kind of previously unseen or undiscovered objects or relations they would correspond to in $\hat{\mathcal{H}}^n$. Several approaches for finding the inverse transformation can be followed, and neural networks are among the obvious choices (Valdés 2002b).

5. An Astronomic Example

In order to illustrate the possibilities of the proposed approach, a dataset containing information about 33055 galaxies was used. This information is part of the Canada-France-Hawaii Legacy Survey (the CFH telescope), and the observational conditions, and preprocessing related with the dataset were the following:

- I-band (red filter) exposure time is 46740 seconds
- Total 5-band exposure time is 77180 seconds (u,g,r,i,z filters were applied, and the i-Band was used to get the morphologies)
- Image reduction, with photometry, and photometric redshifts (courtesy of Stephen Gwyn from the University of Victoria)
- The seeing was 0.9 arcseconds to 1.1 arcseconds (moderate).
- Morphologic analysis by David Schade (Herzberg Institute for Astrophysics, National Research Council Canada)

Each galaxy was characterized by a collection of 11 attributes: *1)* The I-band (red) magnitude, *2-6)* five variables describing the color of the gallaxy (derived from the values obtained by the u,g,r,i,z filters, *7)* the half-light radius of the galaxy image, *8)* the half-light radius, as a measure of the size of the galaxy, *9)* the exponential index of the slope of the light profile, *10)* the axial ratio (longer half-axis/smaller half-axis of an ellipse), and *11)* the Photometric redshift.

When presenting the VR spaces corresponding to the experiments, it must be taken into account that it is impossible to illustrate appropriately the look, feel and immersion of a virtual reality, color, 3D environment within the limits imposed by printed paper. Thus, grey level screen snapshots from the examples are presented only to give a rough idea. The design of the virtual reality spaces was kept simple in terms of the geometries used, (in particular, behaviors were excluded). The snapshots were simplified w.r.t the information included in the corresponding Υs to avoid information overload. The criterium for computing the VR space was to preserve the similarity structure, and the direct transform between the original space and Υ was found by minimizing Sammon error, with ζ_{ij} given by the euclidean distance in Υ and $\delta_{ij} = (1 - \hat{s}_{ij})/\hat{s}_{ij}$, where $\hat{s}_{ij}$ is Gower's similarity (Gower 1973). In all cases, the computed VR space corresponds to **L**-sets extracted from the database containing all of the galaxies.

5.1. Experiment 1: All Galaxies

For this first experiment, all of the variables were used as descriptor attributes, and the **L**-sets were computed at a similarity threshold of 0.85. In addition, the values of the Photometric redshift were used as a classification criterium, and the galaxies were divided into three classes: < 0.5, in the $[0.5, 1)$ interval, and ≥ 1. Accordingly, additional objects were included in the space, namely, transparent membranes wrapping the classes induced by the previously defined partition. The resulting space is shown in Fig. 3. In the left hand side, each element of the **L**-set is represented as a sphere with a radius proportional to the number of objects of the original database represented by the corresponding **L**-object, thus giving an idea of the relative distribution of the elements of the whole database in the VR space. In the right hand side, the elements of the same space are wrapped with semitransparent surfaces corresponding to the classes induced by

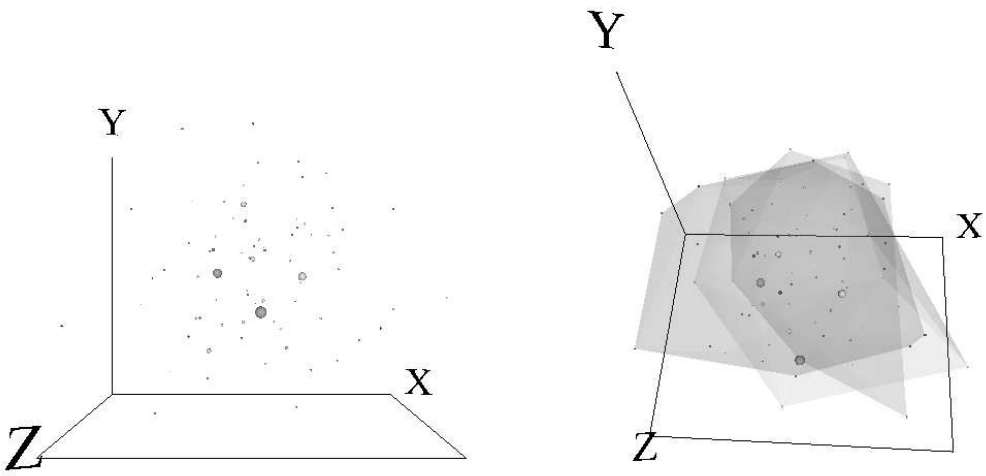

Figure 3. Virtual Reality Space corresponding to the 33055 galaxies database. Left: **L**-set computed with a similarity threshold of 0.85. Right: the same set but with transparent membranes wrapping subsets having specific ranges of the Photometric redshift attribute (see text).

the partition derived from the Photometric redshift. This variable is related with the distance to a given galaxy, and the differential concentration of the galaxies within each class. The clear distinction of the wrapping surfaces, indicates that their intrinsic properties have a dependency w.r.t. their Photometric redshift.

5.2. Experiment 2: Three Groups of Galaxies According to the Photometric Redshift

In this case, the dataset was partitioned into three separate subsets according to the value of the Photometric redshift as described in the previous experiment. Then, a four-fold set of VR-spaces was computed (for the whole dataset, and for the three subsets). In all cases the Photometric redshift was excluded as a descriptor attribute in order not to bias the computation of the **L**-sets and their corresponding VR-spaces, hence, each galaxy was described by a set of 10 attributes. The similarity threshold used for computing the **L**-sets in all cases was 0.75, and the results are shown in Fig. 4.

The shapes and structure of the VR-spaces corresponding to the galaxy subgroups in comparison with the whole are different. This provides an indication of the influence of distance to the galaxy (expressed by the Photometric redshift), on its nature and properties.

6. Acknowledgents

The author would like to thank David Schade from the Herzberg Institute of Astrophysics (for providing the data, and for his suggestions when constructing the VR-spaces), as well as to Alan Barton from the Institute for Information Technology. Both institutes belong to the National Research Council Canada.

7. Conclusion

The construction of virtual reality spaces for astronomic databases allows the visualization and the understanding of the underlying structure of datasets, possibly large. As illustrated by examples from the domain of galaxy research, this tool is potentially useful in knowledge discovery and data mining in astronomy.

References

Borg, I., Lingoes, J. 1987, Multidimensional Similarity Structure Analysis. Springer-Verlag.

Chandon, J.L., Pinson, S. 1981, Analyse Typologique. Thorie et Applications. Masson, Paris.

Fayyad, U., Piatesky-Shapiro, G., Smyth, P. 1996, From Data Mining to Knowledge Discovery. In U.M. Fayyad et al. (eds.) *Advances in Knowledge Discovery and Data Mining*. AAAI Press, pp. 1-34.

Gower, J.C. 1973, A General Coefficient of Similarity and Some of its Properties. *Biometrics* Vol.1 No. 27, pp. 857-871.

Hajek, P., Havranek, T. 1978, Mechanizing Hypothesis Formation. Springer Verlag.

Jianchang, M., Jain, A. 1995, Artificial Neural Networks for Feature Extraction and Multivariate Data Projection. *IEEE Trans. On Neural Networks*. Vol. 6, No. 2, pp. 296-317.

Kruskal, J. 1964, Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psichometrika* Vol 29, pp. 1-27.

Valdés, J.J. 2002, Virtual Reality Representation of Relational Systems and Decision Rules: An exploratory Tool for understanding Data Structure. In Theory and Application of Relational Structures as Knowledge Instruments. Meeting of the COST Action 274 (P. Hajek. Ed). Prague, November 14-16.

Valdés, J.J. 2002b, Similarity-based Heterogeneous Neurons in the Context of General Observational Models. *Neural Network World*. Vol 12., No. 5, pp. 499-508.

Valdés, J.J.,Bonham-Carter G.F. 2003, Virtual Reality Representation of Geoscience Databases and Decision Making Knowledge. Proc. of the 2003 Anual Conference of the Int. Assoc. for Mathematical Geology, Portsmouth, UK, September 7-12.

Valdés, J.J. 2003, Virtual Reality Representation of Information Systems and Decision Rules: An Exploratory Tool for Understanding Data and Knowledge. Lecture Notes in Artificial Intelligence LNAI 2639, pp. 615-618. Springer-Verlag.

Valdés, J.J. 2004, Interpreting fuzzy clustering results with virtual reality-based visual data mining: application to microarray gene expression data. Proc. NAFIPS04 Int. Conf. of the North American Fuzzy Information Processing Society, pp 302-307.

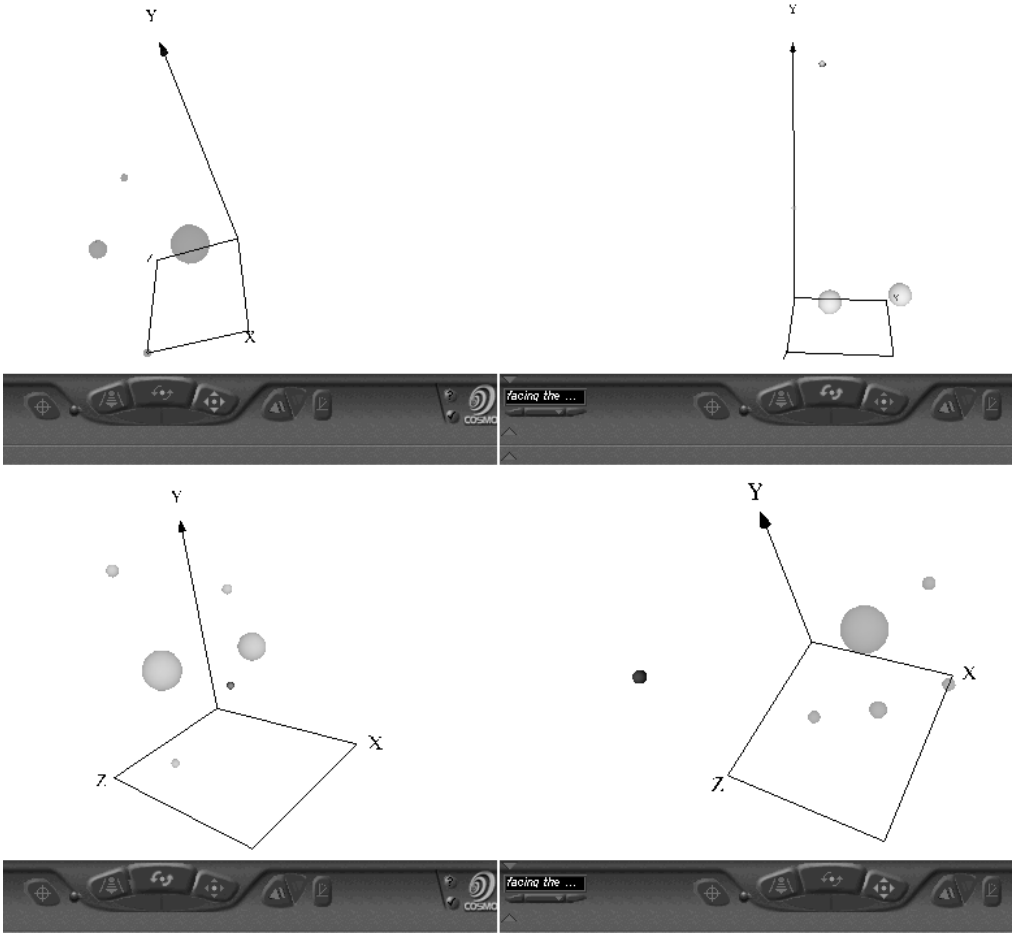

Figure 4. Virtual Reality Space corresponding to the 33055 galaxies database according to the values of the Photometric redshift. The **L**-sets were computed with a similarity threshold of 0.75. Upper left: All of the galaxies. Upper right: Galaxies with Photometric redshift < 0.5. Lower left: Galaxies with Photometric redshift in the $[0.5, 1)$ interval. Lower right: Galaxies with Photometric redshift ≥ 1. The toolbar at the bottom of each representation corresponds to the navigation controls of the virtual reality browser.