Two-Stage Approach to Classifying Multidimensional Cubes for Visualization of Multivariate Data

Hong Thi Nguyen[1], Thuan My Thi Pham[2], Tuyet Anh Thi Nguyen[3], Anh Van Thi Tran[4], Phuoc Vinh Tran[6(✉)], and Dang Van Pham[5]

[1] University of Information Technology, Vietnam National University – HCMC, Ho Chi Minh City, Vietnam
hongnguyen1611@gmail.com
[2] Ho Chi Minh City Open University, Hochiminh City, Vietnam
thuanptm.178i@ou.edu.vn, mythuanpt@gmail.com
[3] Thu Dau Mot University, Binhduong, Vietnam
tuyetnta@tdmu.edu.vn
[4] Ho Chi Minh College of Economics, Hochiminh City, Vietnam
anhttv@kthcm.edu.vn, ttvanh26@gmail.com
[5] Nguyen Tat Thanh University, Hochiminh City, Vietnam
pvdang@ntt.edu.vn
[6] Hochiminh City Open University, Hochiminh City, Vietnam
Phuoc.tvinh@ou.edu.vn, phuoc.gis@gmail.com

Abstract. Visualization of multivariate data is a big challenge to problems of visual analytics. A system of data visualization is composed of visual mapping stage and visual display stage. The stage of visual mapping converts data to graph and the stage of visual display shows the graph on screen in accordance with human's retinal perception which is specified by visual features and Gestalt's laws. Based on data characteristics, multidimensional cubes representing multivariate data are classified as non-spatial multidimensional cube for non-spatial data, spatial multidimensional cube for spatio-temporal data, spatio-temporal multidimensional cube for movement data, and 3D-spatio-temporal multidimensional cube for flight data. For a visualization system responding human's retinal perception, multidimensional cubes have to enable analysts to answer elementary questions concerning individual values, variation questions concerning part of data or overall data, and relation questions resulting in the correlation among attributes.

Keywords: Visualization · Multidimensional cube · Multivariate data Visual analytics

1 Introduction

Visualization of multivariate data is a big challenge to problems of visual analytics for discovering knowledge from data. The problem to be solved is how to represent multivariate data with only one graph displayed on 2-dimensional screen to enable

P. Cong Vinh and V. Alagar (Eds.): ICCASA 2018/ICTCC 2018, LNICST 266, pp. 70–80, 2019.
https://doi.org/10.1007/978-3-030-06152-4_7

analysts to view the whole. Multidimensional cube representing multivariate data is demanded not only to respond data characteristics but also to satisfy human's retinal perception. Visual representation for analytics enables analyst to answer tasks at different levels. Several types of multidimensional cubes enable to answer not only elementary tasks but also questions concerning the variation of an attribute according to reference variable and/or the correlation among attributes.

A system of data visualization comprises two stages, where the stage of visual mapping converts data to visual graph and the stage of visual display transfers the visual graph to human's eyes. The stage of visual mapping responds the characteristics of data, the stage of visual display satisfies the features of human's retinal perception. Based on the basic Gestalt's principle addressed by Kurt Koffka "The whole is more than the sum of its parts" [1, 2], multidimensional cubes are constituted to represent the whole multivariate data on a graph. This study classifies the cubes as non-spatial multidimensional cube for representing non-spatial data, spatial multidimensional cube for spatio-temporal data, spatio-temporal multidimensional cube for movement data, and 3D- spatio-temporal multidimensional cube for flight data.

The paper is structured as follows. The works in the next item are to consider the characteristics of data and the features of human's retinal perception; in that, data are classified as non-spatial, spatio-temporal, movement, and flight; human's retinal perception is considered according to Gestalt's principles, visual features of graph, and demands of visual analytics. Third item considers knowledge discovery from data as the amplification of data corresponding to the increase of human awareness. Fourth item focuses on the taxonomy of multidimensional cubes for representing various types of data. The final item is conclusion.

2 Related Works

2.1 Objects

According to the view of geographic information science, everything existing in real world is depicted as field or object. Field is a mapping from a set of spatial positions onto a set of defined values. Each field is represented as a data variable, of which values distribute extensively out non-boundary space. Object refers to an entity occupying a limited spatial area during a time interval and having boundary determined in space. The existence of an object in real world is depicted by its relation with time or/and space [3, 4]. For example, air temperature exists in space as field, vehicle as object.

Technically, values of field are recorded discretely at different locations, data at other positions are inferred from the recorded values. In geographic information systems, Thiesen polygon and Delaunay triangle are utilized to represent attributes as fields. Data depicting object are classified according to the relation of the object with time or/and space as non-spatial data, spatio-temporal data, movement data, flight data. Space-time cube representing the relation between space and time positions of objects by two axes indicating ground positions and other indicating time is utilized to represent spatio-temporal data and movement data [5].

2.2 Gestalt's Principle

The Gestalt's principle, proposed by Wertheimer, Koffka, and Kohler, refers to the way human perceives images. Retinal perception enables people to understand the signification of image while looking at it. The basic Gestalt's principle "we see the whole before we see the individual parts that make up that whole" is emphasized by Max Wertheimer [6] and other laws were constituted by several authors [1, 2, 6–8]:

- *Figure/Ground* or *Object/Background*: Human's retinal perception has ability to detect an object from background.
- *Area*: For two overlapping objects, the bigger is perceived as background.
- *Similarity*: Objects which have some similar characteristics are perceived as in a group, where similar characteristics may be shape, color, size, and so on. The similarity can be perceived by human vision or experience. Objects of visual connectedness may be perceived as objects of the same group.
- *Continuation*: Human's retinal perception detects objects arranged on the same line or the same curve as a group. Continual arrangement on a line or curve often is detected easier than other features.
- *Closure*: Human's retinal perception has ability to connect discrete objects, e.g. dashes representing trajectory are cognized as a continual line.
- *Proximity*: Objects close to one another are perceived as in the same group.
- *Synchrony*: Synchrony refers to objects of common state, e.g. turned-on lights are the same group, turned-off lights are the same group.
- *Symmetry*: Human's retinal perception tends to form an imagined point or line as symmetric center of objects
- *Parallelism*: Human's retinal perception cognizes parallel objects as a group, e.g. parallel time axes are considered as one.
- *Common region*: Objects located in the same area determined by boundary are perceived as the same group.
- *Past experience*: Individual or public experiences in the past detect objects of the same group.
- *Focal point:* Focal point is an emphasis point to attract viewers.
- *Law of pregnant:* Human's retinal perception tends to convert ambiguous or complex to simple images.
- *Perspective:* Human's retinal capacity perceives that vertical bar is longer than horizontal bar with respect to two equal bars [2]. Mathematically, human's retinal capacity perceives three dimensions of image shown on 2-dimensional display environment.

2.3 Visual Features

Human's retinal perception evaluates the structure of visual graph according to the following features [8–10]:

- *Selection:* Human's retinal perception detects a component of graph or the location of an object on graph.

- *Association:* Human's retinal perception associates the values of a variable or objects of the same characteristic in group.
- *Order:* Human's retinal perception discriminates smaller and larger values of variables, above and below or right and left locations of objects.
- *Quantity:* Human's retinal perception can cognize the ratio of two values which are represented visually.
- *Length:* Length refers to the number of values (elements) of a variable represented on an axis which human can perceive each value.

2.4 Visual Analytics

Visual analytics is an approach to analyzing data based on graph representing visually data. For strategy of visual analytics, analysts' available knowledge and experience are utilized to contribute to the process discovering insights of data. Visualization enables people to easily understand data by using their memory and capacity of imagination [11]. Analyst interacts and views graph representing data to find out its insights by answering questions. Accordingly, the findings depend on not only support of technical tools and way interacting on graph, but also analysts' available knowledge and experience with respect to analytical tasks.

Data analysis refers to answer tasks, where each task is composed of two parts, target and constraint. Target is unknown information, goal of analytical task, where goal or target variable may be reference variable or attribute. Constraint is given information, supposition of analytical task, where constraint or supposition variable may be reference variable or attribute [11]. Bertin classified questions as three levels, elementary, intermediate, and overall. Meanwhile, Andrienko classified questions as two levels, elementary and synoptic. At overall or synoptic levels, if the supposition of question is the combination of space and time variables, there are nine levels of supposition [11]. Generally, a combination of N variables results in 3^N levels of supposition. Questions constituted from triangle of What, Where, When are elementary questions [12, 13].

In our works, non-elementary questions are divided into variation and relation questions. Variation questions refer to an interval or all values of one attribute to consider the change of the attribute over the interval of the attribute [11, 14]. For questions referring to time, the periodic characteristic of time results in the comparison of values during the same period of year or month. Relation questions refer to several attributes sharing a reference variable to consider whether there is correlation between the attributes, e.g. Rainfall and humidity have high correlation with the number of hand-foot-mouth patients, meanwhile temperature does not correlate nearly with the number of hand-foot-mouth patients [15, 16].

3 Two-Stage Approach to Representing Visually Data

3.1 The Process Increasing Human Awareness

Knowledge discovery from data is created by the process amplifying data value. From data of vague significance, information is extracted and analyzed to obtain new knowledge, then generalized to become natural or social laws. The amplification of data value corresponds to the increase of human awareness, from unknown to cognition, then understanding, and generalization. In other words, the significance of data is very hard to be cognized when they have not been processed to become information. Knowledge constituted from findings in information can be generalized to become natural of social laws (Fig. 1).

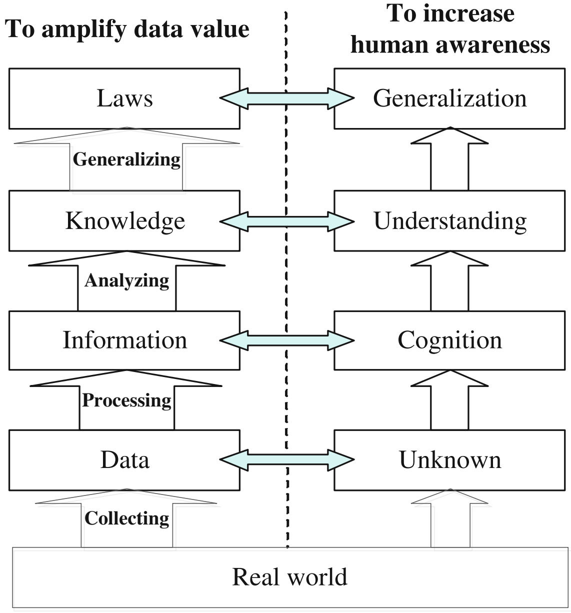

Fig. 1. The process amplifying data value corresponding to the process increasing human awareness.

3.2 Strategy of Knowledge Discovery from Data

For a strategy of knowledge discovery from data, data are converted to information by model or visualization. In a visual analytics system, human and computer collaborate with one another to utilize human's available knowledge and experience in extracting information and discovering knowledge from data by viewing and thinking. Generally, data analytics system combines model with visualization. Technically, a good system of data analytics enables analysts to switch between model and visualization to obtain more insights as possible (Fig. 2) [17, 18].

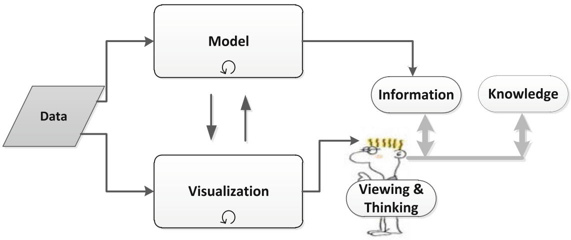

Fig. 2. Data analytics system: data are converted to information and knowledge by model approach and/or visualization approach along with the contribution of human.

3.3 Two-Stage Approach to Representing Visually Data

A basic system of data visualization is composed of two stages, visual mapping and visual display (Fig. 3). For the stage of visual mapping, data are represented as a graph on coordinates; for the stage of visual display, the graph on coordinates is presented and displayed visually on screen. The visual features of graph are improved at the stage of visual display by integrating retinal variables to increase user's retinal perception [9, 15, 19, 20]. Technically, the two stages closely concern one another, some change in a stage affects another. Mathematically, the stage of visual mapping concerns data more than human's retinal perception, the stage of visual display concerns human's retinal perception more than data.

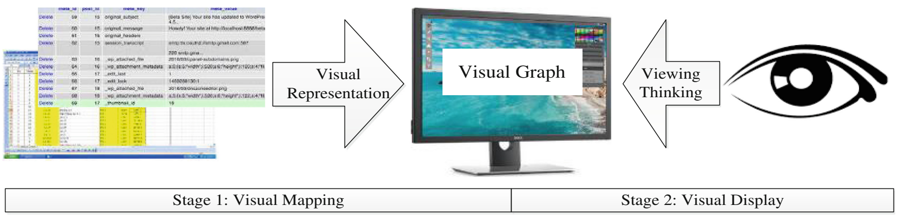

Fig. 3. Two-stage approach to representing visually data.

4 Taxonomy of Multidimensional Cubes

4.1 Non-spatial Multidimensional Cube for Representing Non-spatial Data

Non-spatial multidimensional cube for representing non-spatial data is modified from parallel coordinates, where the common-reference axis is rotated perpendicularly to parallel axes indicating attributes. The variation of attribute is represented on attribute plane formed by the axis of the attribute and the reference axis. Relation plane R may be moved perpendicularly to the reference axis to show the relation between attributes at each reference value or in a reference interval (Fig. 4) [15]. Accordingly, non-spatial multidimensional cube enables analysts to answer elementary, variation, and relation questions in problems of data analytics.

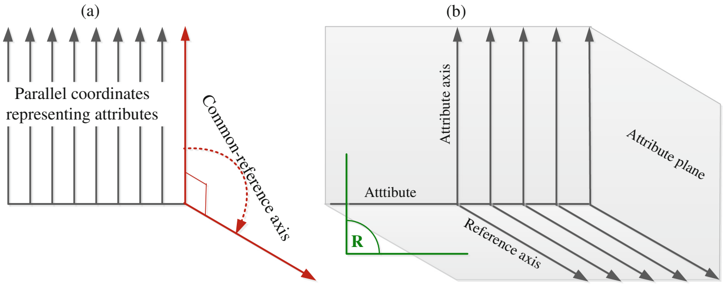

Fig. 4. Non-spatial multidimensional cube for representing non-spatial data: (a) The cube is modified from parallel coordinates; (b) Attribute plane for considering the variation of attribute; Relation plane R for considering the correlation between attributes.

4.2 Spatial Multidimensional Cube for Representing Spatio-Temporal Data

Spatial multidimensional cube, also called multidimensional map, for representing spatio-temporal data is formed by parallel time axes joining perpendicularly to the map at space positions representing data and parallel attribute axes perpendicular to time axes. Each attribute plane formed by attribute axis and time axis shows the variation of the attribute over time. Relation planes perpendicular to time axes show the correlation between attributes at different locations (Fig. 5) [21].

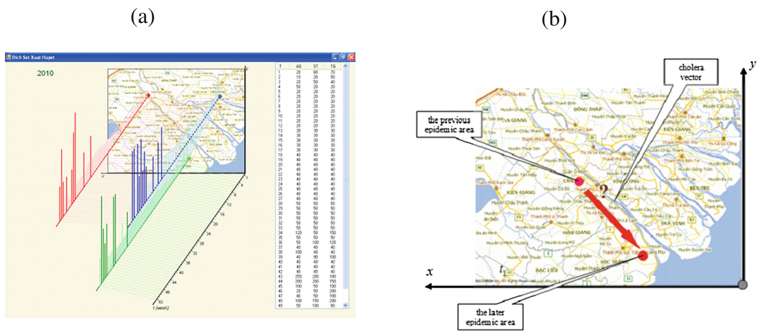

Fig. 5. (a) Spatial multidimensional cube representing the data of dengue fever in Angiang, Soctrang, Tiengiang provinces, Vietnam, during 2009-2012; (b) The topology of epidemic intervals in Angiang and Soctrang is asked whether the epidemic propagates from Angiang to Soctrang, cholera germs propagate along the stream of Mekong river (Source [21]).

4.3 Spatio-Temporal Multidimensional Cube for Representing Movement Data

Spatio-temporal multidimensional cube is constituted by combining a space-time cube [5] with non-spatial multidimensional cube, where the time axis is shared. A moving object changes locations over time, it takes time to move from a location to another, each time position associates one and only one space position, each space position may associate with one or more time positions. Hence, attributes depending on time may be considered as variables referring to time or space. Spatio-temporal multidimensional cube enables to represent the change of attributes while the location of object does not change (Fig. 6).

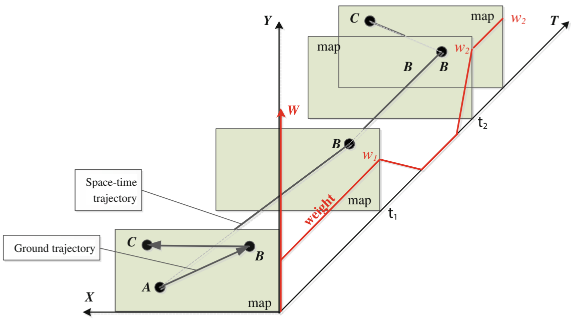

Fig. 6. Spatio-temporal multidimensional cube represents movement data of a lorry moving from A to C passing B. The lorry stops at B from t_1 to t_2 to unload available cargo, its weight decreases from w_1 to 0, then load new cargo, its weight increases from 0 to w_2.

4.4 3D-Spatio-Temporal Multidimensional Cube for Representing Flight Data

3D-Spatio-temporal multidimensional cube for representing flight data combines three cubes, 3D cube, space-time cube, and non-spatial multidimensional cube. In that, 3D cube shares two axes indicating ground positions with space-time cube and space-time cube shares the time axis with non-spatial multidimensional cube. The 3D cube indicates space positions with 2 axes for ground positions and another for elevations; the space-time cube indicates time position; and the non-spatial multidimensional cube indicates attributes referring to time. Each time position on the time axis associates with one and only one space position of flyer on 3D cube; on the contrary, one space position of flyer on 3D cube may associate with one or more time positions on the time axis of space-time cube (Fig. 7) [22, 23].

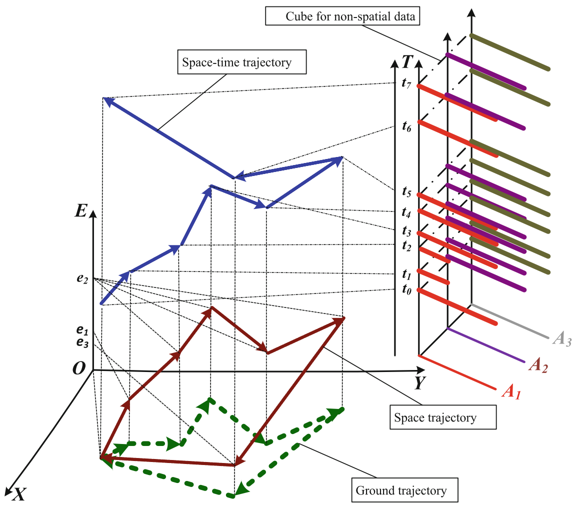

Fig. 7. 3D-Spatio-temporal multidimensional cube for representing flight data combines 3D cube with space-time cube and non-spatial multidimensional cube (Source [22]).

5 Conclusion

The amplification of data value from vague data to knowledge and laws increases human awareness. In the process of data amplification, approaches of model and visualization are utilized separately or simultaneously to extract information from data. Visualization system represents data as visual graph for human's retinal perception. The system is a two-stage approach to represent visually multivariate data, stage of visual mapping and stage of visual display. The stage of visual mapping is affected by data characteristics. The stage of visual display is affected by human's retinal perception specified by visual features and Gestalt's laws.

Approaching data characteristics, this study classifies multidimensional cubes as non-spatial multidimensional cube for representing non-spatial data, spatial multidimensional cube for spatio-temporal data, spatio-temporal multidimensional cube for movement data, and 3D-spatio-temporal multidimensional cube for flight data. Based on analysts' demands, this study also divides analytical questions into elementary questions referring to individual values of variables, variation questions referring to an interval or overall data of a variable, relation questions referring to data of several variables to study the correlation among attributes at different values of reference.

References

1. Tuck, M.: Gestalt Principles Applied in Design (2010)
2. Alexandre, D.S., Tavares, J.M.R.S.: Introduction of human perception in visualization. Int. J. Imaging Robot.™ **4**, 60–70 (2010)
3. Yuan, M., Nara, A., Bothwell, J.: Space–time representation and analytics. Ann. GIS **20**, 1–9 (2014)
4. Yuan, M.: Representing complex geographic phenomena in GIS. Cartography Geographic Inf. Sci. **28**, 83–96 (2001)
5. Hagerstrand, T.: What about people in regional science?. In presented at the Ninth European Congress of Regional Science Association (1970)
6. Bradley, S.: Gestalt Principles: How Are Your Designs Perceived? (2010)
7. Graham, L.: Gestalt theory in interactive media design. J. Hum. Soc. Sci. **2**, 1–12 (2008)
8. Few, S.: Data visualization for human perception. In: Soegaard, M., R.F. Dam, Eds., 2nd ed The Encyclopedia of Human-Computer Interaction. Aarhus, Denmark, 2014
9. Bertin, J.: General theory, from semiology of graphics. In: Dodge, M., Kitchin, R., Perkins, C. (eds.) The Map Reader. Theories of Mapping Practice and Cartographic Representation, pp. 8–16. John Wiley & Sons Ltd, London (2011)
10. Green, M.: Toward a Perceptual Science of Multidimensional Data Visualization: Bertin and Beyond. In: ERGO/GERO Human Factors Science, Citeseer (1998)
11. Andrienko, N., Andrienko, G.: Exploratary Analysis of Spatial and Temporal Data - A Systematic Approach. Springer-Verlag, Berlin Heidelberg (2006)
12. Peuquet, D.J.: Representations of Space and Time. Guilford, New York (2002)
13. Peuquet, D.J.: It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems. Ann. Assoc. Am. Geographers **84**, 441–461 (1994)
14. Aigner, W., Miksch, S., Muller, W., Schumann, H., Tominski, C.: Visualizing time-oriented data—A systematic view. Comput. Graphics **31**, 401–409 (2007)
15. Nguyen, H.T., Tran, A.V.T., Nguyen, T.A.T., Vo, L.T., Tran, P.V.: Multivariate cube integrated retinal variable to visually represent multivariable data. EAI Endorsed Trans. Context-aware Syst. Appl. **4**, 1–8 (2017)
16. Nguyen, H.T., Tran, A.V.T., Nguyen, T.A.T., Vo, L.T., Tran, P.V.: Multivariate cube for representing multivariable data in visual analytics. In: Context-Aware Systems and Applications, Thu Dau Mot, Viet Nam, 2016, pp. 91–100
17. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B.C., Ellis, G., Keim, D.A.: Knowledge generation model for visual analytics. IEEE Trans. Visual Comput. Graphics **20**, 1604–1613 (2014)
18. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G.: Visual analytics: definition, process, and challenges. In: Kerren, A., Stasko, John T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 154–175. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70956-5_7
19. Nguyen, H.T., Ngo, D.N.T., Bui, T.T., Huynh, C.N.T., Tran, P.V.: Visualizing Space-Time Map for Bus. In: Cong Vinh, P., Ha Huy Cuong, N., Vassev, E. (eds.) ICCASA/ICTCC - 2017. LNICST, vol. 217, pp. 38–47. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-77818-1_4
20. Bertin, J.: Semiology of graphics: Diagrams, networks, maps. ed: University of Wisconsin (1983)

21. Tran, P.V., Nguyen, H.T., Tran, T.V.: Approaching multi-dimensional cube for visualization-based epidemic warning system - dengue fever. In: presented at the 8th International Conference on Ubiquitous Information Management and Communication, ACM IMCOM 2014, Siem Reap, Cambodia (2014)
22. Hong, N.T., Phuoc, T.V.: Multidimensional cube for representing flight data in visualization-based system for tracking flyer. In: The 5th International Conference on Control, Automation and Information Sciences, Ansan, Korea, 2016, pp. 132–137
23. Tran, P.V., Nguyen, H.T., Tran, T.V., Bui, T.T.: On an Approach to Visualizing Data of Flying Objects. In presented at the IEEE 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014), Gwangju, South Korea (2014)