다차원 데이터의 3차원 가시화 기법

임강희° 이태동 변성욱 정창성
고려대학교 전자공학과
{ikh, lyadlove, sungwook}@snoopy.korea.ac.kr csjeong@charlie.korea.ac.kr

MRV : 3D Visualization Method for Multidimensional data

Kang-Hee Im° Tae-Dong Lee Sung-Wook Byun Chang-Sung Jeong
Dept. of Electronics Engineering, Korea University

요 약

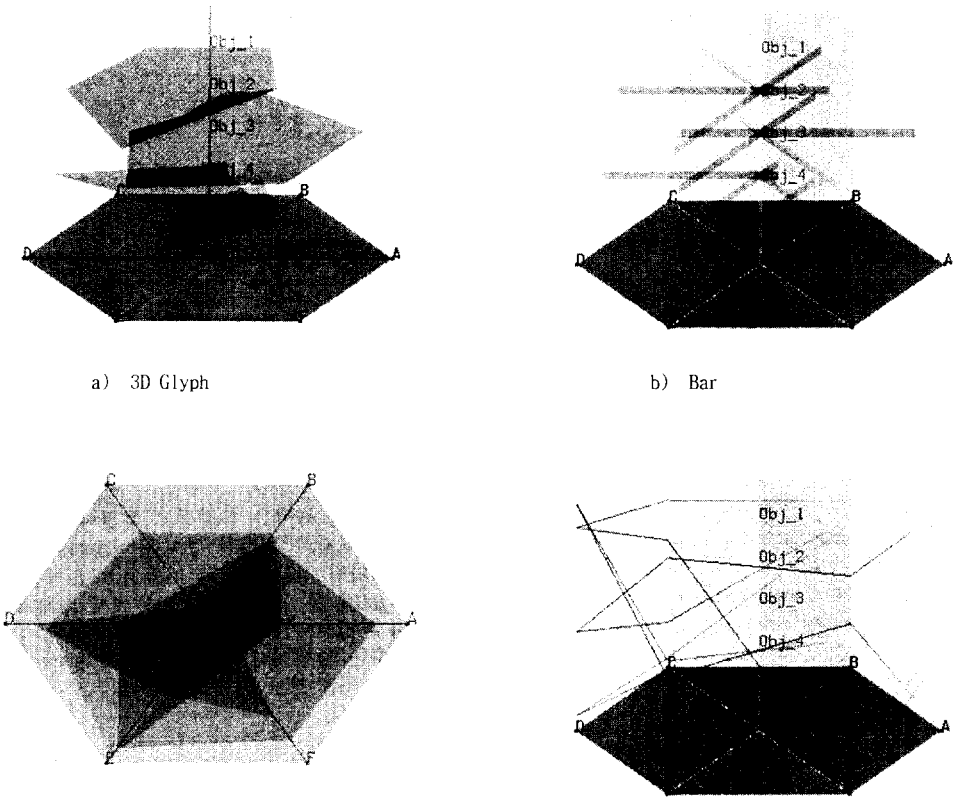
다차원 정보 가시화(multidimensional information visualization)의 목적은 복잡하고 차원이 많은 정보 데이터(information data)를 이해하기 쉽게 그림이나 도표와 같은 특정한 형식을 이용하여 효과적으로 나타내고 비교하는데 있다. 그 동안 제시되어온 다차원 정보 가시화 기법의 대표적인 것으로는 Scatterplots, Perspective Wall, Parallel Coordinates, Glyph를 들 수 있다. 본 논문에서 소개하는 multidimensional rotating visualizer (MRV)이란 기존의 다차원 정보 가시화 기법들을 보완하여 다차원 데이터(multidimensional data)를 3차원 형식으로 보여주는 방법이다. MRV는 그 중에서 특히 Glyph와 Parallel Coordinate의 특징을 혼합하여 화면상에 다차원 정보 데이터를 보여주는 새로운 시도라고 하겠다.

1. 서론

정보 데이터 가시화 기법(information data visualization method)은 크게 계층 정보 가시화(hierarchical information visualization)와 다차원 정보 가시화(multidimensional information visualization)로 나눌 수 있다. 계층 데이터(hierarchical data)에서는 노드(node)와 노드 사이에 문맥(context)이라는 개념으로 데이터들이 관계를 이루고 있는 반면, 다차원 데이터(multidimensional data)에서는 이러한 실재 가시화(visualization)시 직접 표현되는 문맥이라는 개념이 존재하지 않는다. 대신에 차원(dimension)이라는 개념이 있기 때문에 각 차원간의 데이터들은 보이지 않는 연결 관계를 형성하게 된다. 따라서 데이터를 표현할 수 있는 방법이 다양해 질 수 있다. 각각의 가시화 기법은 나름대로의 특징을 띄게 되는데 어떤 방법은 단순 명쾌하여 이해하기가 쉬운 반면 차원간의 비교가 어렵고 어떤 방법은 반대의 경우가 될 수 있는 상호보완(trade-off)관계가 발생하게 된다. 그럼에도 불구하고 지금까지 제시된 다차원 정보 가시화 기법은 특정한 소수의 방법에 그치고 있다. 본 논문에서는 기존의 방법들에 대한 특징을 살펴보고 새로운 방법을 제시해서 함께 비교하고자 한다.

- Scatterplots : 가장 직선적이며 고전적인 방법인 Scatterplots은 수치 데이터의 분석에 좋아서 과학적 가시화(scientific visualization) 등에서 유용하게 쓰일 수 있는 방법이다[1,2]. 전체 화면은 N * N 의 격자(grid)로 세분화되는데 각각의 격자는 두 개의 차원간 비교의 용도로 쓰이게 된다. 이 방법은 결과의 패턴을 통해 두 차원간의 상관관계(정비례, 반비례)의 특성을 파악하기 좋다.
- Parallel Coordinate : 이 방법은 다차원 데이터를 간단한 두 차원간의 표현으로 변형하여 나타내는 방법으로 등간격으로 놓은 수직선들이 각각 하나의 차원에 해당한다[1,3]. 각 차원을 통과하는 선분의 모습을 관찰하여 경향, 패턴, 상관관계를 밝혀 낼 수 있다.
- Glyphs : 가시화기법에서 Glyph는 데이터 값을 그래픽 특성을 지닌 표현으로 나타낸 방법을 뜻한다[1,4]. 일반적으로 Glyph라고 하면 그 중에서도 Star Glyph를 말하는 데 구체적으로 묘사하면 각각의 선들이 공통된 중심을 가지며 같은 각도만큼 차원이 방사형으로 뻗어져 나오게 하는 방법으로, 이 때 그 길이 성분에 데이터 값을 매핑(mapping)시킨 후 그 끝점을 이어 폴리곤(polygon)을 형성하는 것을 말한다.

위와 같이 기존에 제시된 대표적인 다차원 정보 가시화 기법들의 특징들에 대해 알아보았다. 각각의 방법은 나름대로의 장점과 단점을 동시에 지니고 있다. 본 논문에서 새로이 제시하고자 하는 Multidimensional Rotating Visualizer (MRV)는 그 동안의 방법들을 보완, 변형하여 데이터를 3차원 형태로 표현, 분석하려는 시도이다.



a) 3D Glyph b) Bar

c) 2D Glyph d) 3D Coordinate

그림 1. MRV의 다양한 모드

2. MRV Model

MRV는 다차원 데이터를 표현하는 방법의 하나이다. 기본적으로 3차원 형태를 취하고 있는데 그림에서 보듯이 원기둥 형식을 기반으로 하고 있다. 원기둥의 중심을 이용하여 표현하는데 필요한 차원의 수만큼 같은 각도로 분할하여 방사형으로 직선(또는 평면)을 뻗어나가게 한다. 이렇게 뻗어 나온 직선(평면)이 각각의 차원을 할당받게 된다. 기본적으로 중심으로부터 떨어진 거리가 데이터 값 정보를 표현하는 수단이 된다. MRV에서는 모드(mode)가 존재하는데, 3D Glyph 모드, 2D Glyph 모드, Bar 모드, 3D Coordinate 모드로 나누어 표현하고 있다(그림 1). 또한 색상이나 투명도의 개념을 도입하여 이해를 돕는 수단으로 사용한다.

차원의 개수는 기본적으로 정해진 것이 아니고 표현하고자 하는 데이터의 차원을 분석하여 자동적으로 설정된다. 수직 방향을 향하고 있는 중심 축은 기본적으로 표현하고자 하는 분석 주체가 위치하는 곳이다(여기서 분석 주체라 함은 예를 들어, 어떤 집단의 과목별 성적 데이터를 표현할 때 각 개인을 지칭하는 것으로 정의한다). 분석 주체의 개수만큼 중심 축에 등 간격의 기준 좌표가 마련된다. 그리고 그 기준 좌표로부터 독립된 차원 개수만큼의 데이터 값이 방사형으로 길이 성분으로 표현된다.

MRV는 Scatterplot방법처럼 화면을 분할하여 나타내는 방식이 아닌 하나의 화면에 여러 차원을 겹쳐 표현하는 방식을 사용한다. 그리고 3차원 표현법을 사용하기 때문에 뒤쪽에 위치한 차원은 상대적으로 관찰하기 힘들다. 따라서 기본적으로 회전시키는 방법을 사용하고 관찰자 눈의 위치를 상하 좌우로 움직일 수 있어 원하는 위치에서 다차원 데이터를 관찰하게끔 하고 있다.

3. MRV 조작(operations)

3.1 속성 조절(attributes modification)

- 채색(coloring) : 색(color)은 분석 주체간에 서로 다른 값으로 설정되어 있다. 따라서 3차원 상에 데이터가 투영될 때 겹치는 부분이 발생할 수 있는데 서로 다른 색을 이용하여 구별이 용이하도록 하였다. 이런 채색 효과를 선으로 표현하느냐 아니면 면적으로 표현하느냐에 따라 여러 표현법이 나타난다.
- 명암(shading) : MRV에서는 3차원 요소 중에서 특히 투명도(transparency)가 중요한 요소인데 그 이유는 차원이 많아질수록 겹쳐져서 보여지는 부분 또한 증가하기 때문이다. 투명도가 낮으면 뒤에 가린 데이터를 관찰

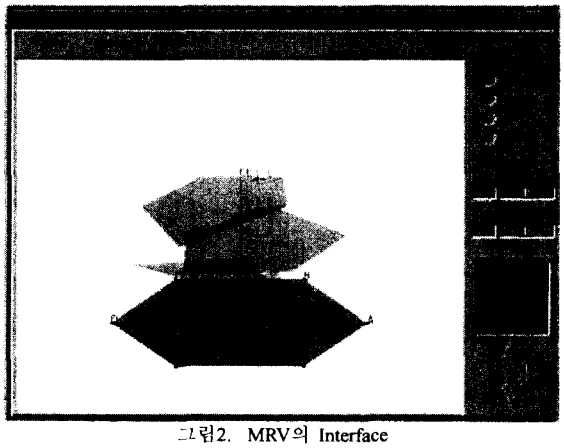


그림2. MRV의 Interface

하기 어렵고 반대로 높으면 지나지게 한 번에 투과되어 보여지는 데이터가 많아져 산만해 질 수 있다.

3.2 표현 방식 변경(geometry modification)

- 3D Glyph : 기존의 Glyph방법은 여러 개의 분석 주체에 관한 MIV를 나타내기 위해 여러 개의 Glyph를 동시에 띄우는 방법을 사용하였지만 여기서는 전반적인 모습을 3차원 상에서 바라볼 수 있게 하였다(그림1-(a)).
- Bar : bar를 이용하여 볼 수 있는 데, 설정 시 중심축을 분석 주체로 잡고 반지름 방향을 각 차원으로 잡았을 경우 하나의 주어진 차원에 대해 분석 주체끼리의 데이터 비교가 용이한 방법이다(그림 1-(b)).
- 2D Glyph : 3D Glyph를 중심 축 위에서 바라보면 2D Glyph를 얻을 수 있다. 위에서 내려다보면 상대적으로 밑에 놓이는 Glyph는 가리게 되지만 투명도 효과를 이용하여 비교가 가능하다(그림 1-(c)).
- 3D Coordinate : 3D Coordinate라는 명칭은 기존의 Parallel Coordinate와 비슷한 면이 있어 붙였다(그림 1-(d)). 중심 축 방향을 따라 차원 수만큼 등 간격의 차원 축이 형성되고 반지름을 따라 분석 주체들이 놓이게 된다.

3.3 대화식(interactive)

- 회전을 통한 관찰 : 다차원 데이터를 3차원 상에서 분석하기 위한 기본 기능이다.
- 차원 위치 변경 : 보여지는 차원의 순서를 임의로 배열할 수 있다.
- 차원 끄고 켜기(turn on, off) : 분석하고자 하는 데이터의 차원의 개수가 많아서 분석이 용이하지 않을 때 원하지 않는 차원을 임시로 끄고 켤 수 있다.

4. 실험

본 실험은 OpenGL 과 Xforms 라이브러리(library)를 사용하여 SGI O_2에서 수행되었다. 처음에 화면에 나타나는 모습은 (그림 2)와 같다. 원하는 위치에서 데이터를 관찰하기 위해서는 상하 좌우 회전을 이용할 수 있다. 데이터가 복잡할 경우 특정 차원 축을 마우스로 클릭해 끄고 켤 수 있다. 아울러 상대적인 차원 배열도 마우스를 이용해 변경이 가능하다. 래디오 버튼을 이용하여 원하는 모드로 바꿀 수 있다. 데이터의 복잡도에 따라 차원 축과 데이터의 투명도를 조절할 수 있다.

5. 평가

MRV의 특징은 차원간의 시각적인 위치 이동을 대화식으로 바꿀 수 있어 여러 차원 중 관심 대상만 효율적으로 비교할 수 있고 또한 차원이 매우 큰 데이터에 대한 추출 기능이 있어 원하지 않는 차원을 끌 수 있다는데 있다. 차원이 증가한다고 해서 Scatterplot처럼 하나 하나의 차원 비교시 작은 화면에 분할하여 표현 하는 방식을 취하지 않고 Parallel Coordinate처럼 중첩해서 표현하는 방식을 취한다. 그러나 MRV는 차원간의 간격을 비교하기 좋게 사용자가 임의로 넓힐 수 있고 잠시 동안 원하지 않는 차원을 끄고 다시 켤 수 있는 융통성을 제공한다. 각각의 모드는 특징이 있는데 Glyph 모드는 전체적인 모습 관찰에 효과적이고, Bar 모드는 어느 특별한 차원을 대상으로 여러 분석 주체간의 비교도 가능하게 해 준다. 2D Glyph 모드는 분석 주체간 정밀 비교가 가능하며 3D Coordinate 모드에서는 기존 Parallel Coordinate의 장점을 살려 비교할 수 있다.

물론 이렇게 하나의 화면에 여러 차원을 동시에 나타내게 되면 겹쳐서 뒤에 배치되는 차원은 가려지는 단점이 있다. 그런 단점은 기본적으로 회전을 통해 관찰자의 위치 변경, 투명도를 조절 기능, 여러 선택 모드 기능으로 보완할 수 있다. 여러 가지 방법으로 바꾸어 가며 관찰하는 동안 데이터에 관한 전체적인 직관력이 생기게 되며 또한 동시에 여러 개의 창을 다른 모드로 띄워 놓고 동시에 효과적으로 비교할 수도 있다.

6. 결론

본 논문에서는 다차원 데이터를 나타내는 새로운 방법으로 MRV를 제시하였다. 이 방법은 기존의 방법이었던 Scatterplot, Parallel Coordinate, Glyph 등을 기반으로 해서 새로운 방식을 제안, 3차원 상으로 확장을 시도하였는데 그 결과 3차원 표현 기법의 장점을 살려서 다양하고 특징 있는 가시화 기법들을 제공할 수 있었다. 앞으로는 모든 정보 표현 기법이 사실적 표현을 극대화시키는 방향으로 나아갈 것으로 예상되므로 딱딱한 정보 표현을 3차원 표현 방식으로 구현하려는 이러한 시도는 중요하다고 하겠다.

7. 참고 문헌

[1]. Allen R. Martin, Matthew O. Ward."High Dimensional Brushing for Interactive Exploration of Multivariate Data" In Proceedings of the 6th IEEE Visualization '95, 1995.

[2]. R. A. Becker and W. S. Cleveland. "Brushing scatterplots", Technometrics, 29(2):127-142, 1987.

[3]. Inselburg, A. Dimsdale, B., "Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry", In Proceedings of the First IEEE Conference on Visualization.,361 ,1990.

[4]. J. H. Siegel, E. J. Farrell, R. M. Goldwyn, and H. P. Friedman. "The surgical implication of physiologic patterns in myocardial infarction shock", Surgery, 72:126-141, 1972.