

정보시각화의 새로운 분류법에 관한 연구

배준우[†] · 이석원 · 김인수 · 명노해

고려대학교 공과대학 산업시스템정보공학과

A Research for New Taxonomy of Information Visualization

Jun-Woo Bae[†] · Suk Won Lee · In-Soo Kim · Rohae Myung

Department of Industrial System and Information Engineering, Korea University

Since too much information has been generated, it became very difficult to find out valuable and necessary information. In order to deal with the problem of information overload, the taxonomy for information visualization techniques has been based upon visualized shapes such as tree map, fisheye view and parallel coordinates, so that it was difficult to choose the right representation technique by data characteristics. Therefore, this study was designed to introduce a new taxonomy for the information visualization by data characteristics which defined by space (3D vs. multi-dimensions), time (continuous vs. discrete), and relations of data (qualitative vs. quantitative). To verify the new taxonomy, forensic data which were generated to investigate the culprit of network security was used. The result showed that the new taxonomy was found to be very efficient and effective to choose the right visualized shape for forensic data for network security. In conclusion, the new taxonomy was proven to be very helpful to choose the right information visualization technique by data characteristics.

Keywords : Information Visualization, Taxonomy, Forensic Data

1. 서 론

1.1 정보시각화

최근 디지털 전자통신 기술의 발전과 정보통신 도구의 증가로 인해 사람들은 매 순간마다 엄청난 양의 정보를 접하게 되는 정보의 홍수 속에서 살게 되었다. 버클리 대학의 연구에 따르면 앞으로 3년 동안 생산될 데이터가 현재까지 만들어 온 정보의 양과 같다고 한다 (Keim, 2002). 이런 엄청난 양의 정보를 일일이 확인하는 것은 사실상 불가능한 일이기 때문에 필요한 정보를 정확히 찾아내는 방법이 중요하게 되었다. 가치 있는

정보가 숨겨져 있을 때 찾아내는 것은 상당히 어려운 일이다. 더욱이 이런 정보가 모두 글자로만 이루어져 있다면 말 그대로 정보의 바다 한가운데서 수영을 하는 것과 같다. 이런 점에서 봤을 때 정보시각화의 목적은 정보를 사람이 보기 쉬운 시각적인 형태로 보여주고 이를 토대로 정보에 대한 통찰력을 가지며 정보와의 상호작용을 통해 의사결정에 도움을 주게 되는 것이다(Keim, 2002).

정보시각화는 독자적인 연구 분야로써는 10년 정도의 역사를 가지고 있지만 최근 급속히 발전하고 있다. 정보 검색, WWW의 하이퍼텍스트, 전자도서관, 인간-컴퓨터 상호작용(HCI) 등 여러 분야에 걸쳐 적용되고 있다.

논문접수일 : 2008년 11월 11일 논문수정일 : 2008년 12월 02일 게재확정일 : 2008년 12월 03일

[†] 교신저자 blackblood@korea.ac.kr

※ 본 연구는 2단계 BK21사업과 한국과학재단을 통해 교육과학기술부의 바이오연구개발사업으로부터 지원받아 수행되었습니다 (M10640030004-08N4003-00410).

정보시각화의 미학은 정보의 상태를 가장 올바르게 나타내는 것이다(Chen, 2006). 현재 이 점은 연구자들이 특정한 정보를 시각화하기 위해 어떤 전략과 도구를 사용할 것이냐를 찾는 형태로 나타나고 있다. 따라서 연구자들이 올바른 시각화 기법을 선택하기 위한 정보 시각화의 분류법이 필요하다(Chen, 2006).

1.2 데이터의 분류

일반적으로 정보는 데이터로부터 생성된다. 대량의 데이터를 다룰 때 우리는 데이터 간의 구조, 특징, 패턴, 경향, 이례적 사항, 또는 관련성들을 파악하려고 한다(황규민, 2004). 이 때 시각화는 다양한 데이터 전체를 정렬하여 보여줌으로써 데이터 전체 또는 데이터 각각에 대한 이해를 도와준다. 즉 사용자가 쉽게 파악할 수 있도록 방대하거나 복잡한 데이터의 집합을 전체적으로 요약하거나 적합한 변수나 관심대상에 따라 분류를 하는 것이다. 데이터는 특정 목적의 주제 또는 응용 분야에 나타낼 수 있는 수치(value) 또는 명칭(name)일 수도 있으며 연속적(continuous)인 성격을 지녔거나 아니면 비연속(discrete) 속성을 가질 수도 있다. 일반적으로 각각의 데이터 요소들은 하나의 기준 값으로 정렬되어 표현되거나, 각 요소간의 상대적 거리 또는 절대 값을 표시하기도 한다. 그 예로 자기공명 단층촬영(MRI) 데이터는 삼차원 공간에서의 데이터 밀도를 보여주고 있으며, 원격 탐사(remote sensing) 데이터는 이차 또는 삼차 공간의 망상구조에서 보여지고, 재무제표 데이터는 기하학적 구조가 아닌 명목 데이터와 수치 데이터의 리스트 형태로 표현된다. 이용자가 보다 쉽게 이해하고 또 그 특징을 파악할 수 있도록 다양한 속성의 데이터를 시각화하는 것은 보편적인 현상이며 각 데이터의 시각화 방법은 다양하다(Grinstein and Ward, 2002). 본 연구에서는 현재까지 알려진 정보 시각화의 방법을 효율적으로 사용할 수 있도록 데이터에 따른 분류 방법에 대해 논의하고자 한다. 다만 2차원 이하의 방법들은 특별한 시각화 방법이 없어도 표현이 간단하므로 3차원 이상을 표현하는 방법들을 대상으로 선정하였다.

2. 기존의 분류법

정보시각화의 방법들을 분류한 기존의 분류법은 대부분 표현 형태를 분류 기준으로 하고 있다. 즉, 결과적으로 어떻게 표현되는지에 따른 분류이다. 새로운 분류법을 제시하기에 앞서 기존의 분류법으로 분류된 시각화 방법 중 대표적인 몇몇을 소개하고자 한다.

2.1 트리(tree) 구조

트리, 혹은 계층제(hierarchies)는 정보의 전체적인 구조를 나타낼 때 흔히 쓰이는 방식이다. 계층제를 나타내는 시각화 방법은 정보시각화의 영역에서 가장 숙성되고 활성화된 분야 중 하나이다(Chen, 2006). 트리 구조를 나타내는 시각화 방법으로는 원뿔 트리(Cone Tree)와 트리 맵(Tree Map)이 대표적이다.

3차원 원뿔 트리(Robertson et al., 1991)는 그림자를 이용한 표현, 투명도, 애니메이션 효과가 상호작용을 하는 형태로 이러한 점은 지금까지도 정보시각화의 원형이 되고 있다(Chen, 2006). 원뿔 트리는 주로 계층제를 나타낼 때 이용되는데 이러한 형태의 계층적 구조의 데이터는 그 자체의 가치를 나타내는 것보다 각 데이터가 다른 어떤 데이터와 어떤 관계를 이루는지가 주제가 된다(Spence, 2007). 이러한 계층적 구조를 나타내는 원뿔 트리는 같은 그룹, 즉 동일한 계층에서 동일한 상위 데이터를 가지고 있는 데이터는 정성적(qualitative)으로 표현된다.

트리 맵 방식의 시각화는 컴퓨터 파일의 계층적 구조를 표현하는 또 하나의 전통적인 방식이다(Johnson and Shneiderman, 1991; Shneiderman, 1992). 트리 맵도 원뿔 트리와 동일하게 계층적 구조에서의 관계를 나타내며 원래는 2차원의 직사각형 공간을 채우는 방식으로 계층적 구조를 나타내었으나 최근에는 3차원 트리 맵도 개발되었다(Bladh et al., 2004). 트리 맵은 표현 방법의 특성상 같은 계층의 데이터를 그 속성에 따라 순차적으로 나타내므로 정성적 관계로 데이터를 표현한다.

2.2 네트워크 구조

네트워크 시각화는 오늘날의 정보시각화의 초석이 되는 분야이다(Chen, 2006). 주위에서 가장 흔하게 볼 수 있는 지하철 노선도도 하나의 네트워크 구조라고 할 수 있다. 복잡하고 다양한 네트워크를 어떻게 시각화하고 축소, 업데이트시키느냐는 정보시각화의 인기 분야이자 도전 과제이다.

NicheWorks(Wills, 1999)는 전화 발신 데이터의 방대한 네트워크를 시각화하는 것을 목적으로 개발된 도구이다. 명목적 관계를 가지는 각 데이터를 하나의 마디(node)로 정하고 각 마디를 연결선(edge)으로 연결해 네트워크를 구성하는 시각화 방법이다. 이 연결선은 네트워크가 어떻게 구성되어 있는지를 나타냄과 동시에 각 데이터가 가지고 있는 연결의 중요도를 연결선의 길이를 통해 나타낸다. 따라서 NicheWorks는 네트워크를 정량적(quantitative)으로 표현하는 시각화 방법이다.

SemNet(Fairchild et al., 1988)은 3차원 시각화의 전통적인 방법으로 지식 구조의 복잡한 연결 관계를 보기 쉽게 나타낸 것이다. 이 방법은 다차원으로 구성된 데이터를 3차원 공간에 나타내며 네트워크의 복잡한 연결을 간략하게, 그리고 정성적으로 나타낸다.

2.3 공간 정보(Spatial Information)의 시각화

공간 정보를 기하학적으로 나타내는 분야는 정보시각화의 가장 중심이 되는 분야이다. 이 분야의 시각화 방법들은 해당 데이터가 공간에서 어떻게 위치하면서 어떤 구간을 형성하고 있는지를 주로 나타낸다.

Bead(Chalmers, 1992)는 정보를 조경도 형태로 나타내는 그래픽 도구이다. 각 데이터를 폴리곤 형태의 노드로 변환시킨 뒤 위에서 내려다보는 형식으로 표현하는 방법으로 데이터의 구조를 정량적으로 나타낸다.

SAGE(Chuah et al., 1995)는 카네기 멜론 대학에서 개발한 시각화 도구로 3차원 공간에서 그래프로 데이터를 표현한다. 이 도구의 특징은 효율적인 시각화를 위해 그래프의 특정 부분을 다른 곳에 자유롭게 옮길 수 있는 것이다. 이 방법은 데이터의 관계를 정량적으로, 또한 시간의 흐름에 따라 나타낼 수 있다.

SPIRE는 PNL(Pacific Northwest National Lab.)에서 개발된 정보시각화 도구로 많은 양의 문서를 읽어야 하는 사용자들을 돕기 위해 개발되었다. 이 도구는 많은 문서를 주제에 따라, 그리고 관계에 따라 3차원 공간에 상징적(metaphor)으로 나타낸다. 또한 이 도구는 시간의 흐름에 따라 변하는 주제와 연결 관계의 변화를 표현한다.

2.4 다변수 데이터(Multivariate Data)의 시각화

다변수 데이터는 여러 개의 독립적인 변수를 가진 데이터이다. 다변수 데이터는 원칙적으로 다차원(multi dimensionality)을 가지기 때문에 이런 데이터의 시각화는 정보시각화 분야의 오랜 도전 과제이다(Spence, 2007).

Parallel Coordinates(Inselberg, 1985, 1997; Wegman, 1990)는 다변수 데이터를 나타내는 가장 인기 있는 방법 중 하나이다. 데이터가 지닌 변수를 횡렬로 평행하게 나열한 다음, 데이터 값을 세로축으로 두고 변수에 따른 데이터 값을 표시하고 선으로 연결시켜 데이터 수만큼 선이 그려지도록 하는 방법이다. 이 방법은 여러 개의 변수를 나타내는 데이터를 정량적으로 표현한다.

Star Plots(Coekin, 1969)는 여러 개의 변수를 가진 하나의 데이터를 보기 쉽게 나타내는 방법이다. 기준 데이터를 중심으로 두고 이와 관련된 여러 변수를 나타내는 데이터를 기준점과 선으로 연결시키고 데이터의

외곽을 연결시킨다. 이 방법도 여러 개의 변수를 나타내는 데이터를 정량적으로 표현한다.

산포도 매트릭스(Scatterplot Matrix)는 다변수 데이터를 위한 산포도로 여러 개의 산포도를 이용해 모든 데이터를 나타내는 방법이다. 이렇게 표현하는 경우 변수가 많아질수록 표현이 힘들어지므로 그래픽 도구를 이용한 산포도 매트릭스(Ahlberg et al., 1992; Ahlberg and Shneiderman, 1994)를 사용해서 많은 변수를 표현할 수 있다. 이 시각화 방법은 기준이 되는 산포도 위에 여러 개의 데이터가 있고 기준을 자유롭게 변형시킬 수 있다. 이 방법은 기본적으로 시간의 흐름에 따른 데이터의 정량적인 관계를 표현한다.

연결 히스토그램(Linked Histogram)은 다변수 데이터를 위한 히스토그램으로 여러 개의 변수를 기준으로 여러 개의 히스토그램을 구성한 뒤 나열하는 방식이다. 이 방법은 그래픽 도구를 이용해서 더 쉽게 사용할 수 있다(Tweedie et al., 1994; Spence and Tweedie, 1998). 이 방법 역시 시간의 흐름에 따른 데이터의 정량적인 관계를 나타낸다.

그리프(Glyph)는 아이콘의 일종으로 명목적, 또는 서열적인 데이터를 대체하기 위한 시각화 방법이다. 그리프는 나타내고자 하는 데이터나 물체와 직접적인 관계가 있거나 없을 수도 있는데, Siirtola(2005)에 의하면 데이터와 연관된 그리프가 13% 더 정확성이 높다고 하였다. 그리프는 정성적인 많은 데이터를 쉽게 나타낼 수 있게 해준다.

Polaris(Stolte et al., 2001)는 다변수 데이터를 다양한 형태로 그래프로 나타내주는 그래픽 도구이다. 이 도구는 시간의 흐름에 따른 데이터의 변화를 나타내며 산포도 매트릭스나 연결 히스토그램처럼 여러 개의 그래프를 통해 여러 변수를 동시에 나타내고 또한 컬러나 형태로 정성적 관계를 나타내기도 한다.

2.5 초점+문맥(Focus+Context) 구조

최근 정보가 매우 많아지면서 표현하고자 하는 정보에 비해 모니터처럼 표현 도구의 크기가 제한되어 있는 경우가 많다. 이런 경우에는 넓은 문맥의 정보를 많이 보여줄 것인가, 또는 좁은 일부의 정보를 자세히 보여줄 것인가의 선택에서 고민하게 되는 경우가 많다. 따라서 자세한 정보를 보면서도 전체적인 문맥을 표현하고자 하는 시도가 많이 이루어져 왔다(Leung and Apperley, 1994). 이런 초점+문맥 구조로 대표적인 것으로 Fisheye View와 Hyperbolic View 등이 있다.

Fisheye View(Furnas, 1986)는 역치(thresholding)를 이용해서 초점을 정하고 이동하면서 정보를 자세히 또는 축

소시켜 보여주는 방법으로 DOI(degree of interest)라는 함수가 이런 왜곡(distortion)의 기준이 된다. 이 시각화 방법은 계층적 구조에서 데이터 간의 연관성과 초점에서의 거리를 고려하여 어떤 정보를 보여주고 삭제할지를 결정하며 정성적 형태의 방대한 계층적 구조나 네트워크 구조의 초점을 강조시켜 알기 쉽게 나타낸다.

Hyperbolic View(Lamping et al., 1995)는 계층별 데이터의 분류를 커다란 하나의 원 안에 나타낸 표현 방법으로 Fisheye View에서 기본 개념을 가져왔다. 이 표현 방법은 계층제와 네트워크 모두를 표현할 수 있고 왜곡을 넣음으로써 더 방대한 데이터를 자세히 보여줄 수 있다. 화면의 초점을 이동함에 따라 구의 중심이 바뀌게 되고 중심에 가까운 정보일수록 상세하게 보여지게 된다. 이 시각화 방법 역시 정성적 형태의 네트워크 구조를 초점에 따라 알기 쉽게 표현한다.

Perspective Wall(Mackinlay, 1991)은 세 개의 패널로 구성된 형태를 가지고 있으며 중앙 패널에는 초점 화면을, 두 개의 측면 패널은 초점 이외의 정보를 축소시켜서 보여준다. 역시 초점이 되는 부분을 정성적으로 나타낸다.

3. 새로운 분류법

3.1 데이터의 분류 기준

앞서 설명한 여러 가지 정보시각화 방법들의 분류는 표현 방법이 주가 되고 있다. 하지만 이런 분류로는 원 데이터를 가지고 있을 때 어떤 방법을 적용해야 효율적인 시각화가 가능한지 명확하지 않다. 따라서 새로운 분류법에서는 어떤 데이터를 표현할지에 따라 다양한 정보시각화 방법들을 분류할 수 있는 것을 목표로 하여 데이터에 따른 효율적인 시각화 방법을 선택할 때 가이드라인이 되고자 한다.

우선 데이터를 어떻게 분류할 것인지를 결정하였다. 정보시각화에서 가장 보편적인 데이터의 분류는 공간에 의해 이루어진다. 그리고 공간은 1차원, 2차원, 3차원, 그리고 다차원으로 이루어진다(Shneiderman, 1996). 이 분류에서 3차원과 다차원을 새 분류법의 분류 기준으로 결정하였다. 1차원과 2차원은 데이터의 단순성으로 인해 특별한 시각화 방법이 없이도 다양한 형태로 표현이 가능하므로 분류에서 제외하였다. 공간은 나타내고자 하는 차원의 수에 따라 3차원과 다차원으로 나누었다. 3차원은 3개의 독립된 축을 가지고 있다는 의미로 이 축들이 이루는 공간을 의미한다. 3차원은 공간을 구성하고 있는 데이터와 3개의 변수로 구성되어 3개의 직교

(orthogonal)하는 축을 가진 데이터도 여기에 포함된다. 하지만 대부분의 디스플레이 도구들은 2차원, 즉 2개의 독립된 축을 가진 평면으로 구성되어 있다. 따라서 시각화 방법을 이용해 하나의 축을 더 표현하는 것이 중요하다. 다차원의 경우는 3차원과 달리 공간을 나타내는 데이터보다는 다변수 데이터, 즉 여러 개의 변수를 포함하고 있는 데이터를 의미한다. 따라서 다차원 데이터는 여러 개의 독립적인 축을 2차원 평면 위에 나타내는 것이 중요하다.

다음으로 각각의 데이터의 관계가 어떤 속성을 가지냐에 따라서도 분류가 이루어진다. 앞서 말한 것처럼 수치를 나타내는 정량적(quantitative)인 데이터가 있고 비수량적(non-quantitative)인 데이터로 순위를 나타내는 서열적(ordinal)인 데이터, 그리고 명칭을 나타내는 명목적(nominal) 데이터가 있다(Mackinlay, 1986). 이 분류에서 정량적인 데이터는 정확한 수치를 가지는 반면 서열적이거나 명목적인 데이터는 정확한 수치적인 차이가 없고 상대적으로 다르다는 점만 가지게 된다. 이름, 모양, 색깔 등이 그 예이다. 따라서 서열적이거나 명목적인 데이터는 시각화에서 동일하게 표현할 수 있고 이에 따라 정성적(qualitative)인 관계로 통일해서 분류하였다. 정량적인 데이터는 정확한 수치를 나타내야 하므로 공간이나 그래프에서 데이터의 위치가 매우 중요하다. 따라서 데이터의 정확한 위치를 모두 나타내기 위해서는 정성적인 데이터에 비해 많은 디스플레이 공간이 필요하다.

마지막으로 데이터가 시간적으로 연속성(continuity)을 가지는가도 분류 기준이 되고 연속성에 따라 연속적(continuous), 비연속적(discrete)으로 나눌 수 있다(Peuquet, 2002). 연속적 데이터는 시간의 흐름에 따라 연속적으로 발생하는 데이터를 말한다. 연속적 데이터를 시각화하기 위해서는 시간의 흐름을 나타낼 수 있는 시각화 방법이 필요하다. 그리고 비연속적인 데이터는 특정한 한 순간에 발생한 데이터이다. 시간의 흐름을 나타낼 필요가 없기 때문에 연속적인 데이터에 비해 표현이 간단하다.

3.2 새로운 분류 기준

이렇게 3가지 분류에 따라 데이터를 분류한 뒤 문헌 연구에서 소개한 정보시각화 방법들이 어떤 데이터를 효과적으로 표현하는지를 분석하여 데이터의 분류 기준에 따라 배치하였다<표 1>.

3차원이면서 정성적 관계를 가진 비연속적인 데이터를 나타내는 방법은 매우 다양하다. 트리, 즉 계층적 구조를 가진 데이터나 네트워크 구조를 가진 데이터는 대부분이 이 분류에 속하게 된다. 계층적 구조는 기본적으로 원뿔 트리나 트리 맵을 적용할 수 있는데 원뿔 트

〈표 1〉 데이터 분류 기준에 따른 특성

데이터 분류 기준 시각화방법	3차원	다차원	비연속적	연속적	정성적	정량적
3차원 원뿔트리	○		○		○	
트리 맵	○		○		○	
NicheWorks	○		○			○
SemNet	○		○			○
Bead	○		○			○
SAGE	○			○		○
SPIRE	○			○	○	
Parallel Coordinates		○	○			○
Star Plots		○	○			○
산포도 매트릭스		○		○		○
연결 히스토그램		○		○		○
그리프		○	○		○	
Polaris		○		○	○	
Fisheye View	○		○		○	
Hyperbolic View	○		○		○	
Perspective Wall	○		○		○	

리의 경우는 한 화면에 모든 데이터와 전체적인 구조를 모두 나타낼 수 있다. 하지만 데이터가 많을 경우 각 데이터가 겹쳐져서 보이기 때문에 원하는 데이터를 따로 찾아봐야 된다. 트리 맵은 디스플레이 크기의 한계로 인해 한 화면에 모든 데이터를 나타내기는 힘들어서 상위 데이터나 최하위 데이터를 빼고 표현할 때도 있다. 따라서 원뿔 트리처럼 전체적인 구조를 한 눈에 알아볼 수는 없다. 하지만 트리 맵은 데이터의 세부적인 속성을 더 쉽게 표현이 가능하다. 따라서 계층적 구조를 전체적인 형태로 나타낼 때는 원뿔 트리가 더 적합하고 계층적 구조 내의 각 데이터의 세부적인 속성을 알기 쉽게 나타낼 때는 트리 맵이 더 적합하다.

Fisheye View나 Hyperbolic View는 방대한 데이터를 가진 계층적 구조나 네트워크 구조를 쉽게 나타내는 것을 목적으로 하고 있다. 이 경우 왜곡이 없이 한 화면에 표현을 하면 너무 작아져서 볼 수 없을 경우에 많이 쓰이는데 Fisheye View는 화면의 여러 데이터 중 중요한 데이터를 함수에 의해 선택하여 확대하고 나머지는 축소시키는 반면, Hyperbolic View는 전체적인 구조에서 현재 초점을 중심으로 해서 초점에서 멀어질수록 축소시키는 형태를 취하고 있다. 따라서 방대한 계층적 구조나 네트워크 구조에서 중요한 특정 데이터를 강조하고 싶을 때는 Fisheye View가, 데이터의 전체적인 구

조를 중심으로 나타내고 싶을 때는 Hyperbolic View가 더 적합하다. 그리고 Perspective Wall은 앞서의 두 방법과 달리 계층적 구조나 네트워크 구조의 데이터가 아니라도 표현이 가능하다. 역시 한 화면에 나타낼 수 없는 대량의 데이터를 나타내지만 특정 데이터를 확대, 축소 시키기보다는 초점과 그 주위를 제외한 나머지 부분을 축소시키는 형태를 띄고 있고 Fisheye View나 Hyperbolic View처럼 특정 데이터를 각각 확대 또는 축소시키는 데는 적합하지 않다.

3차원 정량적 관계를 가진 비연속적인 데이터를 시각화하는 방법으로는 NicheWorks와 SemNet, 그리고 Bead가 효과적이다. NicheWorks는 네트워크 구조를 정량적인 형태로 나타낸 것으로 대량의 네트워크 데이터를 한 화면에 모두 나타낼 수 있다는 장점을 가지고 있다. 그리고 연결선의 길이로 중요한 연결을 눈에 잘 띄게 표현할 수 있지만 하나하나의 데이터를 자세히 볼 때는 적합하지 않다. SemNet 역시 네트워크 구조를 연결선 길이를 통해 정량적인 형태로 나타내는데 대량의 데이터를 한 눈에 알아볼 수 있도록 나타내지 않고 현재 초점이 맞춰진 데이터를 자세하게 보여준다. Bead는 데이터의 종류에 상관없이 3차원 공간 자체를 조경도 형태로 표현하여 대량의 데이터를 표현 가능하고 전체 데이터를 한 화면에 나타낼 경우에는 자세히 볼 수 없지만 데이터를 개별적으로 자세히 검색할 수 있다.

3차원이면서 연속적인 데이터를 나타내는 방법으로는 SPIRE와 SAGE를 들 수 있다. SPIRE는 데이터의 구조를 3차원 공간에 은유적으로, 그리고 시간의 흐름에 따라 나타낸다. SPIRE는 데이터의 구조를 관련성에 따라 그룹으로 표현하므로 정성적 관계를 가진 데이터의 표현에 적합하다. 그리고 SAGE는 데이터를 입체적인 막대 그래프 같은 형태로 나타낸다. 따라서 3개의 변수를 가진 정량적 데이터를 표현하는 데 적합하다.

다차원이면서 비연속적인 데이터는 연속적인 데이터에 비해 비교적 간단히 나타낼 수 있다. 정성적 데이터인 경우 그리프를 이용해 직관적이면서 간단하게 표현할 수 있다. 이 때, 그리프가 해당 데이터의 성격을 잘 표현하는 것이 중요하다. 정량적인 데이터의 경우에는 각 차원의 변수를 축으로 나타내서 수치적인 정보를 보여주는 것이 중요하다. 이때는 여러 변수의 축을 한 번에 나타낼 수 있는 Parallel Coordinates와 Star Plots이 효과적이다. Parallel Coordinates는 다변수 데이터에서 기준이 되는 변수별로 절대적 수치 변화를 알아볼 때와 수치 변화의 전반적인 경향을 나타낼 때 유용하다. 그리고 Star Plots는 기준 변수별 상대적 수치 변화를 알아볼 때나 여러 데이터의 수치 변화의 차이를 알아볼 때 유

용하다.

다차원이면서 연속적인 데이터는 비연속적인 데이터에 비해 표현이 쉽지 않아서 그래픽 도구의 도움을 받는 경우가 많다. 정성적인 관계를 가진 데이터를 시각화하는 도구로는 Polaris가 적합하다. Polaris는 시간의 흐름에 따른 데이터의 변화를 아이콘의 변화를 통해 보여준다. Polaris는 정량적인 데이터를 표현하는 것도 가능하지만 이 경우에는 산포도 매트릭스와 연결 히스토그램이 더 효율적인 시각화를 보여준다. 정량적인 관계를 가진 데이터의 경우에는 산포도 매트릭스와 연결 히스토그램을 쓸 수 있는데 산포도 매트릭스는 각 변수에 해당하는 모든 산포도를 나타내며 변수가 많을 경우 일부 변수만 표현하거나 팝업 형식의 그래픽 효과를 이용하여 정량적인 다차원 데이터를 효과적으로 표현하며 여러 가지 변수를 기준으로 변수의 절대적인 값과 분포를 알아볼 때 유용하다. 그리고 연결 히스토그램의 경우도 산포도 매트릭스처럼 각 변수별로 히스토그램을 나타내어 변수별 데이터를 보기에 효과적이며 상대적인 수치의 차이와 변화량을 나타낼 때 유용하다.

이렇게 데이터의 분류 기준에 따라 <표 2>와 같은 새로운 분류법을 완성하였다.

<표 2> 새로운 분류법

공간	데이터의 관계	시 간	
		비연속적	연속적
3 차원	정성적	원뿔 트리 트리 맵 Fisheye View Hyperbolic View Perspective Wall	SPIRE
	정량적	NicheWorks SemNet Bead	SAGE
다 차원	정성적	그리프	Polaris
	정량적	Parallel- Coordinates Star Plots	산포도- 매트릭스 연결- 히스토그램

4. 포렌식 데이터에 대한 새 분류법 적용

기존의 시각화 분류법은 표현되는 형태 자체가 분류 기준이 되었었다. 그로 인해 데이터의 성격에 맞는 시각화 방법을 찾기가 쉽지 않았다. 그래서 공간, 시간, 그리고 데이터 간의 관계를 기준으로 새로운 분류법을 개발하였다. 이러한 새로운 분류법의 타당성을 검증하

기 위하여 최근 중요성이 커지고 있는 포렌식 데이터에 새 분류법을 적용해 보았다.

포렌식이란 수사의 근본적인 특성에 따라 법적으로 유효한 전자적 증거의 확보를 목표로 과학적 지식과 기술을 활용하여 전자적 증거를 수집하고 분석하는 전반적인 행위를 의미한다(전상덕 등, 2006). 최근에는 개인용 컴퓨터의 하드 디스크가 100GB를 상회하고 네트워크에서 연간 생성되는 데이터의 단위가 EB(Exabyte)에 이르는 만큼 포렌식 데이터의 양도 대단히 많다. 따라서 사이버 수사관들이 방대한 양의 포렌식 데이터를 기본적인 형태로 일일이 확인하는 것은 사실상 불가능하다. 이런 점에서 사이버 수사관들이 효율적으로 조사할 수 있도록 포렌식 데이터를 나타낼 수 있는 시각화가 점점 중요해지는 추세이다(조선영, 2006).

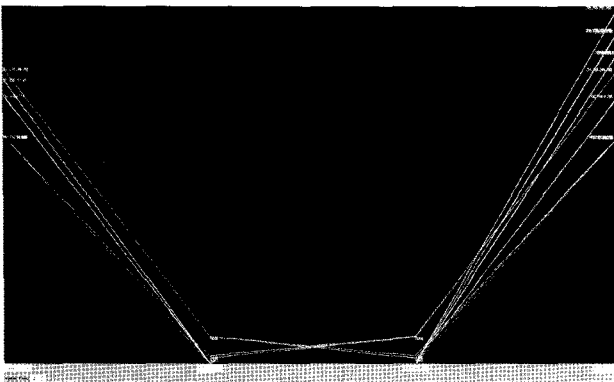
포렌식 데이터는 컴퓨터에 관련된 모든 정보에 대한 조사를 하는 데이터이기 때문에 종류가 다양하다. 그 중 네트워크 데이터는 인터넷 사용이 활발해지면서 네트워크를 통한 불법 침입이 늘어나고 있다는 점에서 매우 중요해지고 있는 추세이다. 특정 컴퓨터에서 다른 컴퓨터에 접속을 하게 되면 여기에 관련된 정보가 생성되는데 여기서 생성되는 데이터로는 발신지(source) IP 주소, TCP 발신지 포트(port), TCP 목적지 포트, 그리고 목적지(destination) IP주소 등이 있다(Conti, 2007). 이 정보를 패킷 데이터라고 하며 이 데이터는 사용하고 있는 컴퓨터에서 어떤 프로그램과 포트를 통해 다른 컴퓨터와 연결되어 있는지를 나타내고, 불법적인 네트워크 연결이 구성되어 있는지를 확인할 수 있는 기본 정보가 된다. 따라서 패킷 데이터는 네트워크 감시를 할 때 꼭 필요한 데이터이다. 이 패킷 데이터의 경우 4개의 변수를 가진 다차원 데이터가 된다. <표 3>는 실제 특정 컴퓨터에서 생긴 패킷 데이터의 일부이다. 하지만 이런 형태로 데이터를 나타내면 어떤 데이터가 비정상적이거나 불법적인 데이터인지 한 눈에 파악하기 힘들고 이런 현상은 데이터가 많아질수록 더 심해지게 된다. 따라서 시각화 방법을 적용해 비정상 데이터를 찾아내기 쉽게 나타내야 한다.

<표 2>에서 제시한 새로운 분류법에 의하면 다차원 이면서 정량적 관계를 가진 비연속적인 데이터인 패킷 데이터를 표현할 수 있는 가장 적합한 방법은 Parallel Coordinates나 Star Plots이다. 두 시각화 방법 중에서 Parallel Coordinates 방법을 Georgia Institute of Technology에서 개발한 Rumint라는 그래픽 도구를 이용해 나타내 보았다<그림 1>. 이렇게 Parallel Coordinates를 이용해 시각화하게 되면 많은 패킷 데이터를 한 화면에 나타낼 수 있고 일반적인 데이터의 형태와 다른, 즉 수상한 형

태의 데이터를 한 눈에 알아볼 수 있기 때문에 텍스트로 나타낸 것보다 훨씬 효율적으로 네트워크 침입을 찾아볼 수 있다.

<표 3> 패킷 데이터

발신지 IP	발신지 포트	목적지 포트	목적지 IP
163.152.30.20	2152	80	209.85.143.147
163.152.30.20	2148	80	58.227.194.78
163.152.30.20	2100	80	209.85.175.147
163.152.30.20	1345	4661	193.42.213.100
163.152.30.20	1291	4500	89.248.172.90
163.152.30.20	1041	5004	211.234.240.252
163.152.30.20	2119	80	211.115.11.222
163.152.30.20	1368	80	121.125.77.147
127.0.0.1	1414	1413	127.0.0.1
127.0.0.1	1406	1407	127.0.0.1



<그림 1> 패킷 데이터의 Parallel Coordinates

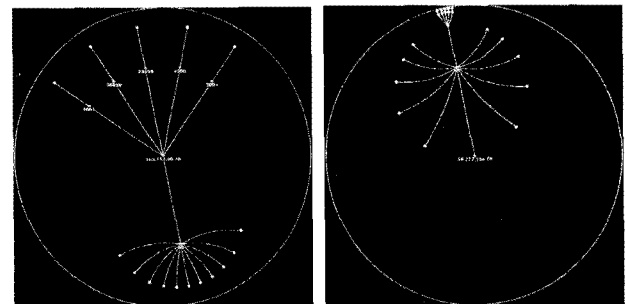
<표 2>에서 분류한 데로 데이터의 분류에 따라 다차원 데이터를 적합한 방법으로 분류할 수 있었다. 이번에는 네트워크 연결 상태의 구조를 나타내는 데이터를 시각화하고자 한다. 이 데이터는 주로 네트워크 관리자가 네트워크 연결 상태를 확인할 때 많이 사용된다. 이러한 경우는 특정 위치에 있는 컴퓨터가 어떠한 다른 컴퓨터들에 연결되어 있는지 나타내는 것이 주제가 되며 이러한 데이터 사이의 관계는 정성적으로 볼 수 있다. <표 4>는 163.152.30.20이라는 주소를 가진 컴퓨터에서 생긴 네트워크 데이터의 일부분이다. 이런 텍스트 데이터로는 네트워크의 전체적인 구조를 파악하기 힘들고 데이터가 많은 경우 일일이 검색을 해야 하는 불편함이 있다. 따라서 한 눈에 네트워크 구조를 파악할 수 있는 시각화가 필요하다. 이 경우 구조 데이터이므로 3차원이고 정성적 관계를 가지며 비연속적인 데이터이므로 새로운 분류법을 적용하면 원뿔 트리를 비롯하여 여

러 가지 방법으로 시각화할 수 있다. <그림 2>는 그 여러 가지 방법 중 Hyperbolic View를 자바 스크립트를 이용해 나타낸 결과이다. <그림 2>의 좌측은 계층적 구조의 중심에 되는 루트(root) 마디가 초점일 때를 나타낸 것이며 우측은 초점을 최하위 마디인 목적지 IP까지 이동했을 경우를 나타낸 것이다. 이 경우 네트워크 데이터의 양이 방대해질 경우에도 연결 구조가 어떤 식으로 이루어졌는지를 한 눈에 파악할 수 있게 된다.

새로운 분류법을 사용하여 포렌식 데이터의 시각화를 실험해 본 결과 데이터의 특성을 이용한 정보시각화가 적합하다고 보여진다.

<표 4> 네트워크 데이터

연결 포트	연결 네트워크
4661	193.42.213.100
36818	71.7.236.172
23695	70.174.136.148
4500	89.248.172.90
5004	211.234.240.252
80	121.125.77.147
	211.115.11.222
	209.85.175.147
	124.137.96.24
	58.227.194.78



<그림 2> 계층적 구조의 Hyperbolic View

5. 결론

지금까지 정보시각화에서 많이 쓰이는 방법들과 적용되는 데이터를 공간, 시간 그리고 데이터 간의 관계에 따른 기준으로 분류해 보았다. 앞서 언급한 것처럼 기존의 정보시각화 방법들의 분류는 표현 방법에 따라 나누었는데 새로운 분류법은 시각화를 적용시킬 데이터의 특성에 따라 정보시각화 방법들이 분류가 되기 때문에 더 편리한 시각화 적용이 가능하게 되었다. 그러므로

정보시각화를 위하여 데이터의 세 가지 특성, 즉 공간적 특성(3차원 vs. 다차원), 시간적 특성(연속적 vs. 비연속적), 그리고 관계적 특성(정성적 vs. 정량적)으로 분류한 새로운 분류법이 원 데이터를 시각화하는데 유용하고 효과적인 방법임이 증명되었다.

참고문헌

- [1] 전상덕, 홍동숙, 한기준; “디지털 포렌식의 기술 동향과 전망”, 정보화정책, 13(4) : 3-19, 2006.
- [2] 조선영; “포렌식 수사 관점에서의 디스크 정보의 효율적인 표현에 관한 연구”, 석사학위논문, 고려대학교, 2006.
- [3] 황규민; “Fisheye view를 이용한 PDA의 메뉴 디자인에 대한 정보 시각화 방안”, 석사학위논문, 고려대학교, 2004.
- [4] Ahlberg, C., and Shneiderman, B.; “Visual Information Seeking : Tight Coupling of Dynamic Query Filters with Starfield Displays,” ACM, Proceedings CHI 1994 : 313-317, 1994.
- [5] Ahlberg, C., Williamson, C., and Shneiderman, B.; “Dynamic Queries for Information Exploration : An Implementation and evaluation,” ACM, Proceedings CHI 1992 : 619-626, 1992.
- [6] Beard, K., Deese, H., and Pettgrew, N. R.; “A Framework for Visualization and Exploration of Events,” Information Visualization, 7 : 133-151, 2008.
- [7] Bladh, T., Carr, D. A., and Scholl, J.; “Extending Tree-Maps to Three Dimensions : A Comparative Study,” APCHI 2004, Proceedings of the 6th Asia-Pacific Conference on Computer-Human Interaction (APCHI 2004), Rotorua, New Zealand : 50-59, 2004
- [8] Chalmers, M.; “Bead : Explorations in Information Visualisation,” Proceedings of SIGIR 1992, Copenhagen, Denmark : 330 - 337, 1992.
- [9] Chen, C.; “Visualising Semantic Spaces and Author Co-Citation Networks in Digital Libraries,” *Information Processing and Management*, 35(2) : 401-420, 1999.
- [10] Chen, C.; Information Visualization : Beyond the Horizon, 2nd Edition, Springer-Verlag, London : 1-142, 2006.
- [11] Chuah, M. C., Roth, S. F., Kolojechick, J., Mattis, J., and Juarez, O.; “SageBook : Searching Data-Graphics by Content,” Proceedings of CHI 1995, Denver, CO. : 338-345, 1995.
- [12] Coekin, J. A.; “A Versatile Presentation of Parameters for Rapid Recognition of Total State,” IEE, Proceedings of International Symposium on Man-Machine Interaction, Cambridge, England, 1969.
- [13] Conti, G.; Security Data Visualization, No Starch Press, Inc., San Francisco : 1-227, 2007.
- [14] Fairchild, K., Poltrock, S., and Furnas, G.; “SemNet: Three-Dimensional Graphic Representations of Large Knowledge Bases,” Cognitive Science and its Applications for Human-Computer Interaction, Guidon, R. (Ed.), Lawrence, Erlbaum : 201-233, 1988.
- [15] Furnas, G. W.; “A Fisheye Follow-up : Further Reflection on Focus+Context,” Proceedings of the Computer Human Interaction, Montreal, Canada : 999-1008, 2006.
- [16] Grinstein, G. G. and Ward, M. O.; “Introduction to Data Visualization,” Information Visualization in Data Mining and Knowledge Discovery, Fayyad U. M., Grinstein G. G., and Wierse A. (Ed.), Morgan Kaufmann Publisher, San Francisco : 21-45, 2002.
- [17] Inselberg, A.; “The Plane with Parallel Coordinates,” *The Visual Computer*, 1 : 69-91, 1985.
- [18] Inselberg, A.; “Multidimensional Detective,” IEEE, Proceedings of Information Visualization 1997 : 100-107, 1997.
- [19] Johnson, B. and Shneiderman, B.; “Tree-maps : A Space Filling Approach to the Visualization of Hierarchical Information Structures,” *Proceedings of IEEE Visualization*, 91 : 284-291, 1991.
- [20] Keim, D; “Information Visualization and Visual Data Mining,” *IEEE Transactions on Visualization and Computer Graphics*, 8 : 1-8, 2002.
- [21] Lamping, J. and Rao, R.; “The Hyperbolic Browser : A Focus Plus Context Technique for Visualizing Large Hierarchies,” *Journal of Visual languages and Computing*, 7(1) : 33-55, 1996.
- [22] Lamping, J., Rao, R., and Pirolli, P.; “A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies,” Proceedings of the ACM CHI'95 Conference on Humna Factors in Computer Systems : 401-408, 1995.
- [23] Leung, Y. K. and Apperley, M. D.; “A Review and Taxonomy of Distortion-Oriented Presentation Techniques,” ACM Transactions on Computer-Human Interaction, 1(2) : 126-160, 1994.
- [24] Mackinlay, J.; “Automating the Design of Graphical Presentations of Relational Information,” *ACM Tran-*

- sactions of Graphics*, 5(2) : 110-141, 1986.
- [25] Mackinlay, J. D., Robertson, G. G., and Card, S. K.; "The Perspective Wall: Detail and Context Smoothly, Integrated," Proceedings of ACM CHI 1991, New Orleans, LA, New York, ACM Press : 173-179, 1991.
- [26] Melanic, T. and Torsten, M.; "Human Factors in Visualization Research," *IEEE Transaction on Visualization and Computer Graphics*, 10(1) : 72-84, 2004.
- [27] Robertson, G. G., Mackinlay, J. D., and Card, S. K.; "Cone trees : Animated 3D Visualizations of Hierarchical Information," Proceedings of CHI 1991, New Orleans, LA : 189-194, 1991.
- [28] Shneiderman, B.; "Tree Visualization with Tree-Maps : A 2-d Space Filling Approach," *ACM Transactions on Graphics*, 11(1) : 92-99, 1992.
- [29] Shneiderman, B.; "The Eyes Have It : A Task by Data Type Taxonomy of Information Visualization," Proceedings of IEEE Workshop on Visual Language, Boulder, CO., : 336-343, 1996.
- [30] Siirtola, H.; "The Effect of Data-Relatedness in Interactive Glyphs," IEEE, Proceedings IV05 : 869-876, 2005.
- [31] Spence, R.; *Information Visualization : Design for Interaction*, 2nd Edition, ACM Press, Inc., Harlow : 1-120, 2007.
- [32] Spence, R. and Tweedie, L.; "The Attribute Explorer : Information Synthesis via Exploration," *Interacting with Computers*, 11 : 137-146, 1998.
- [33] Stolte, C., Tang, D., and Hanrahan, P.; "Polaris : A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases," *Transactions on Visualization and Computer Graphics*, 2001.
- [34] Tuftte, E. R.; *The Visual Display of Quantitative Information*, 2nd Edition, Graphics Press, Cheshire, Conn., : 1-190, 2001.
- [35] Tweedie, L., Spence, R., Williams, D. M. L., and Bhogal, R.; "The Attribute Explorer," ACM, Conference Companion Proceedings CHI 1994, : 435-436, 1994.
- [36] Ware, C.; *Information Visualization : Perception for Design*, 2nd Edition, Morgan Kaufmann, San Francisco, : 4-28, 2004.
- [37] Wegman, E. J.; "Hyperdimensional Data Analysis Using Parallel Coordinates," *Journal of the American Statistical Association*, 85(411) : 664-675, 1990.
- [38] Wills G. J.; "NicheWorks-Interactive Visualization of Very Large Graphs," *Journal of Computational and Graphical Statistics*, 8(2) : 190-212, 1999.
- [39] Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., and Crow, V.; "Visualizing the Nonvisual: Spatial Analysis and Interaction with Information from Text Documents," Proceedings of IEEE Symposium on Information Visualization, Atlanta, Georgia, USA, : 51-58, 1995.