

Visualization of three-dimensional data with virtual reality

Jae Eun Lee^a · Sojin Ahn^a · Dae-Heung Jang^{a,1}

^aDepartment of Statistics, Pukyong National University

(Received February 13, 2017; Revised April 12, 2017; Accepted May 04, 2017)

Abstract

Various data visualization methods are utilized to analyze a huge amount of data. Among various methods, a three-dimensional image requires the rotation of the image to show a stereo image on a two-dimensional screen. This study discusses two methods of batch method and real-time method, which make it possible to construct of stereo images to improve the restriction of the three-dimensional image display with virtual reality. This investigation can be useful to better explore a three-dimensional data structure.

Keywords: data visualization, virtual reality, stereo image, R package

1. 서론

통계학계의 피카소로 불리는 존 튜키는 “사람은 보는 것만으로도 데이터에 대하여 많은 것을 알 수 있으며 데이터의 의미를 이해하고 그 이야기를 이해하곤 한다.”라고 하며 데이터 시각화의 중요성에 대하여 강조하였다. 그만큼 정보의 의미를 빠르고 정확하게 전달할 수 있는 가장 효과적인 방법은 그림 혹은 이미지를 제공하는 것이다. 데이터 시각화란 데이터 분석 결과를 이해할 수 있도록 시각적으로 표현하고 전달하는 과정을 말한다. 때로는 한 장의 그림이 책 한 권의 설명보다 더 설득력 있다는 말이 있다. 이처럼 대용량 데이터들을 이해하기 쉽고 효과적으로 전달하기 위하여 다양한 시각화 방법들을 이용하고 있는데 그 중에서 가장 많이 사용되는 형태가 그래프이다.

3차원 그래프의 경우 물체를 여러 가지 각도로 보아야만 공간적인 입체의 모양을 파악할 수 있다. 하지만 우리가 사용하는 2차원인 모니터 평면 화면으로 3차원 정보를 명확히 표현하고 살펴본다는 것은 매우 어렵다. 따라서 3차원 그림을 살펴보는데 어려움이 있기 때문에 3차원 그림을 한 눈에 살펴볼 수 있도록 돕는 시각화 방법론적 요소가 필요하다. 이에 대한 해결책으로 가상현실(virtual reality; VR)을 이용한 방법론을 제시하고자 한다. 가상현실이란 어떤 특정한 환경이나 상황을 컴퓨터로 만들어서, 실제 주변 상황과 상호작용을 하고 있는 것처럼 만들어주는 것을 뜻한다. 1980년대에 등장한 가상현실은 군사 활용을 시작으로 일반인이 사용할 수 있는 가상현실 게임기를 만들었지만 기술부족으로 인하여 대중화에 실패하였다. 하지만 2010년대 이후 다시 붐이 일어나면서 게임 및 영상 산업에 많이 활용되어 왔다. 이 시기에 대중들이 사용할 수 있도록 만들어진 VR 기기들은 가격이 굉장히 비싸고 약 500g의

This work was supported by a Research Grant of Pukyong National University (Year 2016).

¹Corresponding author: Department of Statistics, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Korea. E-mail: dhjang@pknu.ac.kr



Figure 2.1. Smartphone VR display, VR glasses and VR controller.

헤드셋을 머리에 착용해야하기 때문에 무겁고 거추장스러웠다. 하지만 삼성, 구글, 페이스북 및 다양한 중소기업들이 VR 산업에 본격적으로 투자하고 다양한 제품들을 선보이면서 VR 기기들은 예전보다 훨씬 저렴해지고 가벼워졌으며 고성능 렌즈, OLED 패널 등을 사용하여 선명하게 3차원 그림을 분석할 수 있게 되었다. 또한 거의 모든 사람들이 스마트폰을 사용하고 있기 때문에 스마트폰과 값싼 VR 기기를 결합한다면 많은 사람들이 3차원 그림을 효과적으로 분석할 수 있게 될 것이다.

본 논문의 구성은 다음과 같다. 제 2절은 VR을 이용하여 3차원 시각화 자료를 살펴보는 방법으로 Batch method와 Real-time method에 대한 개념을 설명하고 여러 가지 예시들을 살펴보았다. 제 3절에서는 결론으로 마무리하였다.

2. VR을 이용한 시각화

Figure 2.1은 스마트폰 VR 디스플레이, VR 안경과 VR 컨트롤러이다. 스마트폰을 VR 디스플레이 또는 VR 안경에 장착하여 3차원 그림을 살펴볼 수 있다. VR 기기는 PC에 바로 연결하여 사용하는 일체형 VR 기기와 스마트폰을 장착하여 사용하는 모바일 VR 기기로 나뉜다. 일체형 VR 기기의 경우 고성능 PC가 필요하기 때문에 굉장히 비싸지만 모바일 VR 기기의 경우 수십 가지 이상의 기기들이 있으며 저렴하게 이용할 수 있다. 그리고 VR 컨트롤러는 3차원 시각화 자료를 살펴보는 방법인 Batch method에서 그림을 넘길 때 손쉽게 이용할 수 있다. VR 기기를 이용하여 3차원 그림을 살펴보는 방법은 크게 두 가지로 나눌 수 있다. 첫 번째는 정지된 그림을 살펴보는 것이고, 두 번째는 실시간으로 움직이는 그림을 관찰하는 것이다. 또한 VR을 이용하여 2차원 그림을 3차원으로 살펴보기 위해서는 양안의 좌·우 눈에 비출 두 개의 그림이 필요하며 두 그림의 각도는 입체감을 느낄 수 있게 서로 다르게 설정 해주어야 한다. 이렇게 만들어진 그림을 스테레오 이미지라고 한다. VR을 이용한 3차원 이미지를 살펴보기 위하여 R에서 제공하고 있는 라이브러리 `aplpack` (Wolf, 2015), `rpanel` (Bowman, 2015), `lattice`

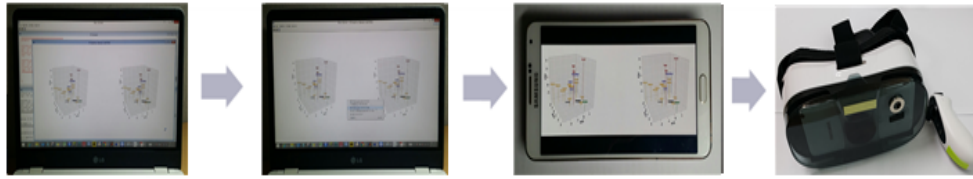


Figure 2.2. The procedure of Batch method for iris data.

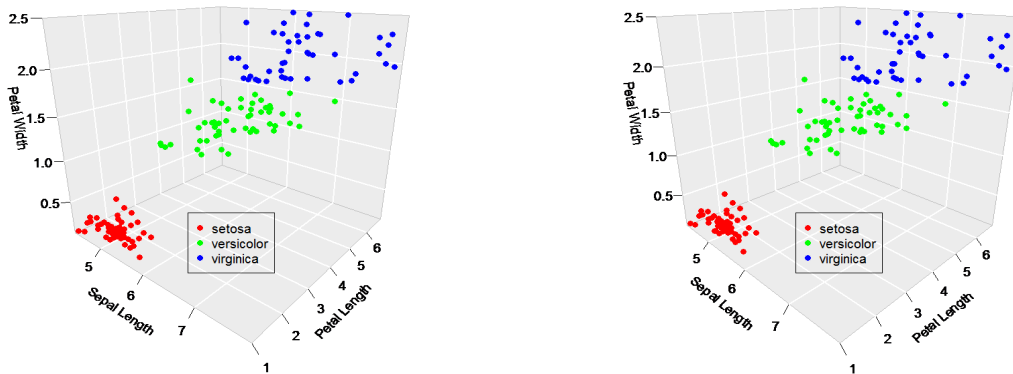


Figure 2.3. The stereo image for iris data.

(Sarkar, 2016), MASS (Ripley, 2016), plot3D (Soetaert, 2016), rgl (Murdoch, 2017), scatterplot3d (Ligges, 2017) 등을 이용하였다.

2.1. Batch method

Batch method는 스마트폰에 저장된 스테레오 이미지를 VR을 이용하여 살펴보는 방법이다. Batch method에 대한 구체적인 절차는 아래의 예제를 통하여 소개하고자 한다. 첫 번째 예제는 5개의 변수(꽃받침 길이, 꽃받침 폭, 꽃잎 길이, 꽃잎 폭, iris 꽃 종류)와 150개 관측 값을 담고 있는 iris 데이터를 이용한 것이다. Figure 2.3은 꽃잎 폭, 꽃잎 길이, 꽃받침 길이를 이용하여 iris 3종류의 꽃(setosa, versicolor, virginica)에 대하여 그린 3차원 산점도에 대한 스테레오 이미지이다. Batch method를 이용하여 수행하는 과정은 다음과 같으며 Figure 2.2를 통하여 구체적인 절차를 확인할 수 있다.

단계 1: R 프로그램을 이용하여 Iris에 대한 데이터를 분석하고 스테레오 이미지를 생성한다.

단계 2: 생성된 Iris 스테레오 이미지를 컴퓨터에 저장한다.

단계 3: 컴퓨터에 저장한 Iris 스테레오 이미지를 스마트폰으로 다운 받은 후 저장한다.

단계 4: VR 안경 또는 VR 디스플레이와 VR 컨트롤러를 이용하여 스마트폰에 저장된 Iris 스테레오 이미지를 확인한다.

Batch method를 이용하여 3개의 집락 및 점들의 위치를 입체적으로 확인할 수 있다. 이 작업을 위한 R 코드는 부록 A.2에 주어져 있다.

두 번째 예제는 미국 펜실베이니아 주립대학교 ‘Center for Astrostatistic’ 사이트 (<http://astrostatistics.psu.edu>)에서 제공한 은하계 자료(The Sloan Digital Sky Survey (SDSS) galaxy data)로서 23개의 변

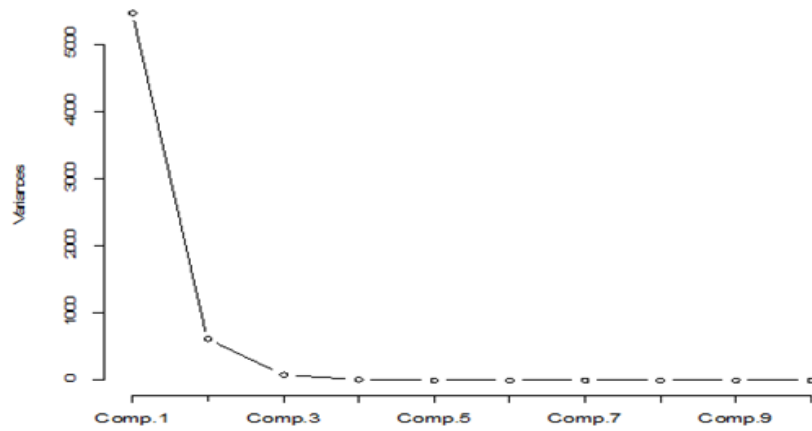


Figure 2.4. The scree graph for galaxy data.

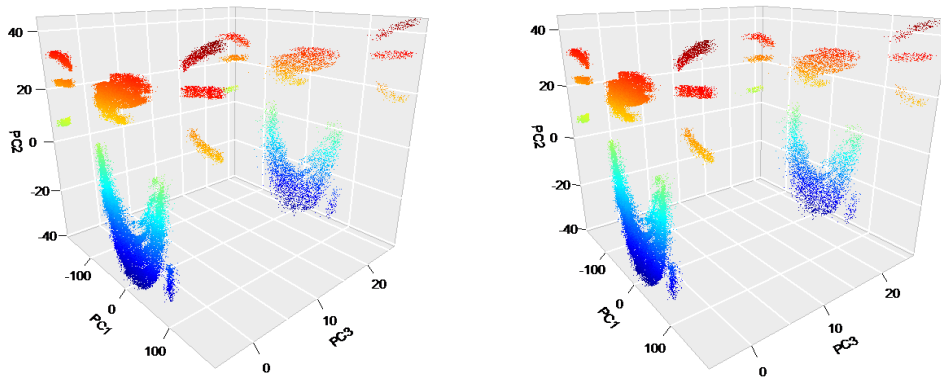


Figure 2.5. The stereo image of principal component score plot for galaxy data.

수와 약 46,420개의 준항성(quasar)의 관측값을 이용하여 주성분 분석을 한 것이다. 은하계 데이터는 23개의 변수들로 이루어져 있으며 항성과 은하들에 대한 정보를 가지고 있다. 주성분 개수를 선택하기 위하여 누적 기여율을 구한 결과 주성분이 두 개인 경우 98.1%, 세 개인 경우 99.6%의 변동에 대한 설명비율을 보여주고 있다. 또한 Figure 2.4의 스크리 그래프를 살펴보면 두 개의 고유값이 가파른 경사 위에 있고 나머지 고유값들은 완만한 경사 위에 있다. 이 경우 제1주성분과 제2주성분인 두 개의 주성분으로 축약하여도 자료 설명이 충분하지만 제3주성분까지 고려한 결과 Figure 2.5처럼 특이한 구조의 3차원 그림이 그려지게 된다. Figure 2.5의 스테레오 이미지를 컴퓨터에 저장하여 스마트폰으로 다운받고 저장한 후 VR 안경과 VR 컨트롤러를 이용하여 점들의 위치를 입체적으로 확인할 수 있다. 이 3차원 그림에서 그려진 제3주성분 축의 작은 값 영역에는 대다수의 점들로 구성된 특이한 기하학적 구조를 갖는 집락들이 나타나고 제3주성분 축의 큰 값 영역에는 소수의 점들로 구성된 같은 구조를 갖는 집락들이 나타난다.

세 번째 예제는 22개의 미국 공공시설에 대한 전기사용 자료(electric utility data)를 이용하여 주성분 분석을 한 것이다. 이 자료는 Oracle Data Mining and Analytics 사이트(<http://oracledmt.blogspot.kr>)

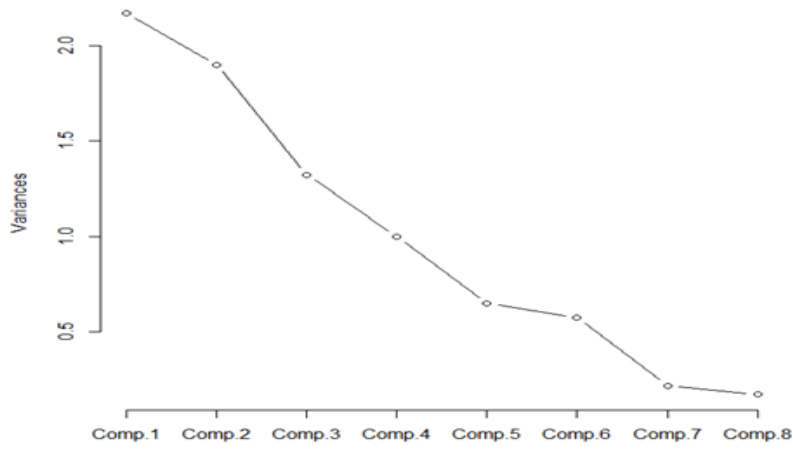


Figure 2.6. The scree graph for electric utilities.

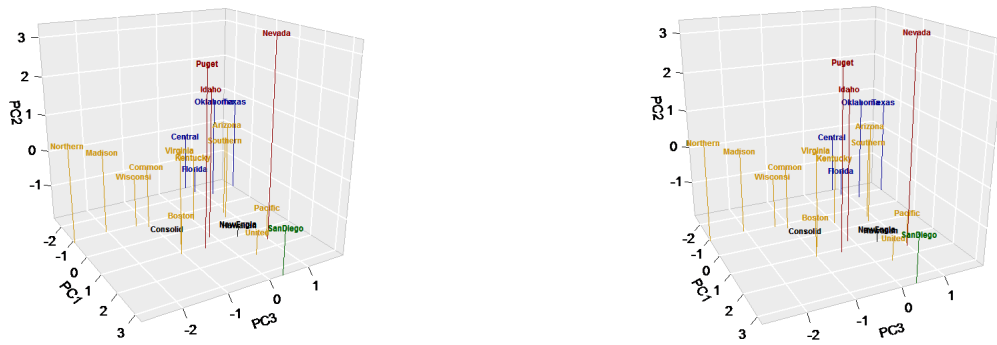


Figure 2.7. The stereo image of principal component score plot for electric utilities.

/2007/06/way-cooler-pca-and-visualization-linear.html) (Campos, 2007)에서 수집하였다. 22개의 관측값과 8개의 변수(고정 요금 적용 비율(소득/부채), 자본 수익률, KW 용량 당 비용, 연간 하중 계수, 1974-1975년 KWH 최고 수요 증가, 판매(KWH 연간 사용), 핵 퍼센트, 총 연료 비용(KWH당 센트))로 구성되어 있다. 첫 번째와 두 번째 주성분을 합친 누적 기여율은 약 50%이고 세 번째 주성분까지의 누적 기여율은 약 70%이다. Figure 2.6인 스크리 그래프를 살펴보면 세 개의 고유값이 가파른 경사 위에 있다. 일반적으로 전체 분산의 70-90%가 되도록 주성분의 수를 결정하기 때문에 세 개의 주성분을 이용하여 분석하였다. 두 번째 예제에서는 제2주성분과 제3주성분의 누적 기여율이 거의 같기 때문에 제1주성분과 제2주성분을 이용했을 때와 제3주성분까지 이용했을 때의 그림 구조가 매우 비슷하게 나타난다. 반면 세 번째 예제에서는 제1주성분과 제2주성분을 이용했을 때와 제3주성분까지 이용했을 때 다른 구조의 형태가 그려지며 이에 따라 새로운 정보를 파악할 수 있다. Figure 2.7은 제3주성분까지 이용하여 그린 삼차원 주성분 점수 그림에 대한 스테레오 이미지이다. 이 스테레오 이미지를 컴퓨터에 저장하여 스마트폰으로 다운받고 저장한 후 VR 안경과 VR 컨트롤러를 이용하여 각 관측값 간 거리 및 각 집락 간 거리를 입체적으로 비교할 수 있다.

Table 2.1. Central composite design with three responses

Natural Variables		Coded Variables		Responses		
ζ_1	ζ_2	x_1	x_2	y_1	y_2	y_3
80	170	-1	-1	76.5	62	2940
80	180	-1	1	77.0	60	3470
90	170	1	-1	78.0	66	3680
90	180	1	1	79.5	59	3890
85	175	0	0	79.9	72	3480
85	175	0	0	80.3	69	3200
85	175	0	0	80.0	68	3410
85	175	0	0	79.7	70	3290
85	175	0	0	79.8	71	3500
92.70	175	1.414	0	78.4	68	3360
77.93	175	-1.414	0	75.6	71	3320
85	182.07	0	1.414	78.5	58	3630
85	167.93	0	-1.414	77.0	57	3150

$x_1 = \text{time}$; $x_2 = \text{temperature}$; $y_1 = \text{yield}$; $y_2 = \text{viscosity}$; $y_3 = \text{molecular weight}$.

Oracle Data Mining and Analytics에서는 미국 공공시설의 전기사용에 대하여 주성분 3개를 이용하여 삼각형 그림(triangular (or ternary) graphs)을 그렸으며 22개의 공공시설들 중 비슷한 공공시설끼리 5개의 집락으로 나누었다. 여기서 삼각형 그림은 색도도(chromaticity diagram)를 이용하여 3차원 직각좌표 공간에서의 점들을 삼각형 형태의 2차원 심플렉스 공간에서의 점으로 표현한 것이다. 그러나 3개의 주성분들은 서로 직교하기 때문에 3차원 직각 좌표계를 이용하여 3차원 산점도를 작성하여야 하는데 삼각형 형태의 2차원 심플렉스 공간에서의 산점도로 그림을 그리면 각 관측값 간 거리 및 각 집락 간 거리에 왜곡이 발생한다. 반면 Figure 2.7은 서로 독립인 주성분 축들로 그려진 직각 3차원 그래프를 그려 각 관측값들 사이의 거리를 정확하게 표현하고 있다.

네 번째 예제는 Myers 등 (2009) 책에 나오는 세 개의 반응변수 값이 있는 화학처리에 대하여 중심합성 계획법(central composite design)을 이용한 것이다. 이에 대한 자료는 Table 2.1에서 볼 수 있다. 변수들은 두 개의 공정변수인 시간(x_1), 온도(x_2)와 세 개의 반응변수인 수율(yield; y_1), 점성(viscosity; y_2) 그리고 분자량(molecular weight; y_3)이 있다. 각 반응변수들에 대한 추정값을 계산하면 아래와 같은 모형으로 나타낼 수 있다.

$$\begin{aligned}
 \hat{y}_1 &= 79.94 + 0.99x_1 + 0.52x_2 - 1.38x_1^2 - 1.00x_2^2 + 0.25x_1x_2, \\
 \hat{y}_2 &= 70.00 - 0.16x_1 - 0.95x_2 - 0.69x_1^2 - 6.69x_2^2 - 1.25x_1x_2, \\
 \hat{y}_3 &= 3386.2 + 205.1x_1 + 177.4x_2.
 \end{aligned} \tag{2.1}$$

각 반응변수마다 최적화 해가 있으며 반응변수가 많을 경우 동시에 최적화가 되는 해를 결정하기 위하여 우리는 주로 비선형 계획법(nonlinear programming methods)이나 소망도함수(desirable function) 기법을 사용한다. 각 반응변수에 제약조건이 있는 최적화 문제(constrained optimization problem)를 아래의 식과 같이 나타낼 수 있다.

$$\text{Max } y_1 \text{ subject to } 62 \leq y_2 \leq 68, y_3 \leq 3400. \tag{2.2}$$

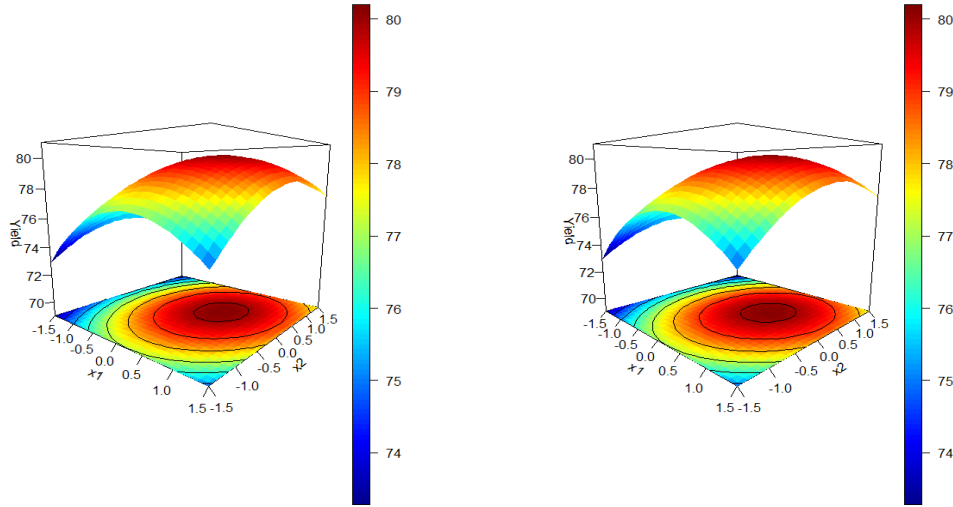


Figure 2.8. The stereo image for the estimated response equation for the first response, yield.

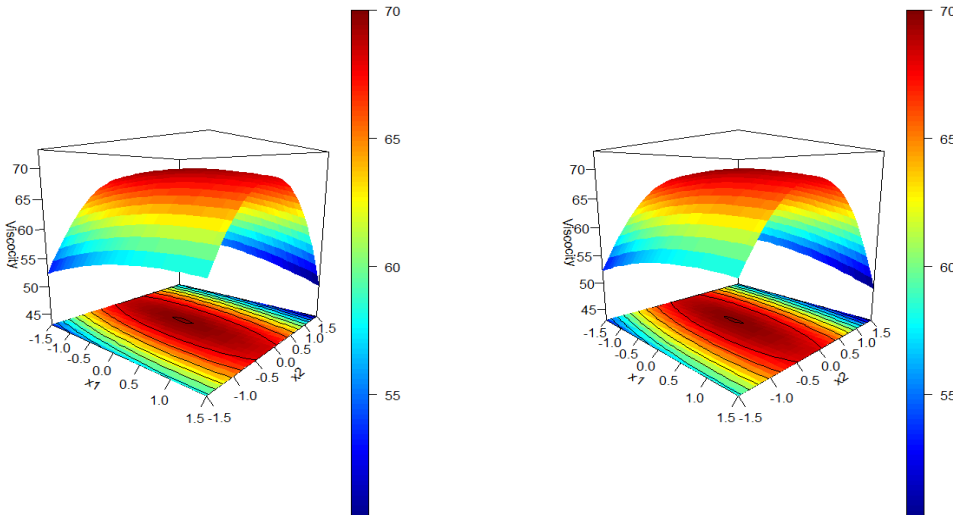


Figure 2.9. The stereo image for the estimated response equation for the second response, viscosity.

이러한 최적화 기법을 적용하기 전에 하는 탐색적 자료 분석 작업이 Figures 2.8–2.10 및 Figures 2.11–2.13과 같이 각 반응표면에 대한 그림들을 그리는 것이다. R 패키지를 이용하여 추정된 세 개의 모형에 대한 그림들을 그려보면 Figures 2.8–2.10 및 Figures 2.11–2.13처럼 3차원 표면도 및 등고선도를 그릴 수 있으며 VR을 이용하여 각 반응변수들에 대한 반응표면을 입체적으로 확인 할 수 있다. Figures 2.11–2.13은 식 (2.2)를 만족하는 부분($62 \leq y_2 \leq 68, y_3 \leq 3400$)에는 반응표면을 그리고 만족하지 않는 부분에는 회색으로 그린 후 y_1 에 대한 반응표면 위에서 탐색적으로 찾은 두 개의 최적해를 표시한 그림이다. 이렇게 찾은 두 개의 최적해는 비선형 계획법으로 찾은 두 개의 최적해와 일치한다. Figures 2.11–2.13에서 보듯이 3차원 그림을 2차원으로 볼 경우 두 최적화 해의 위치를 색을 통하여 대략적으로

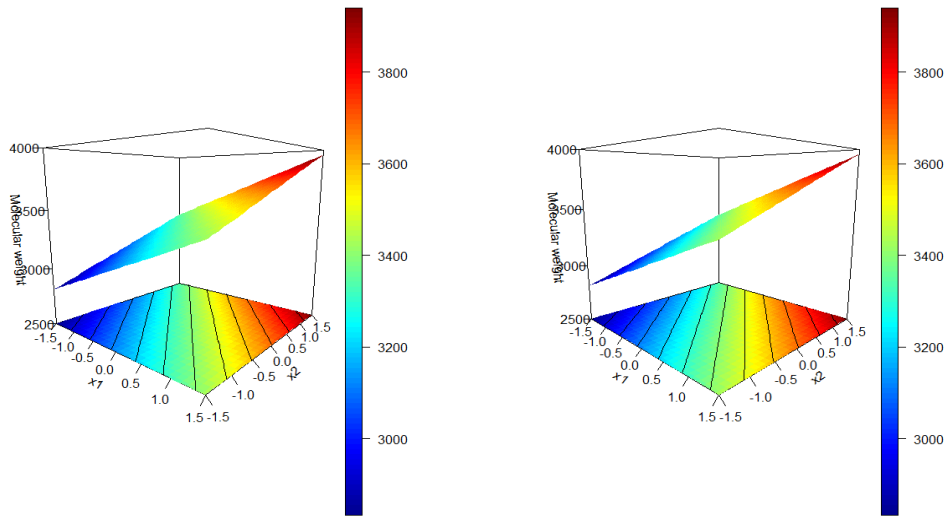


Figure 2.10. The stereo image for the estimated response equation for the third response, molecular weight.

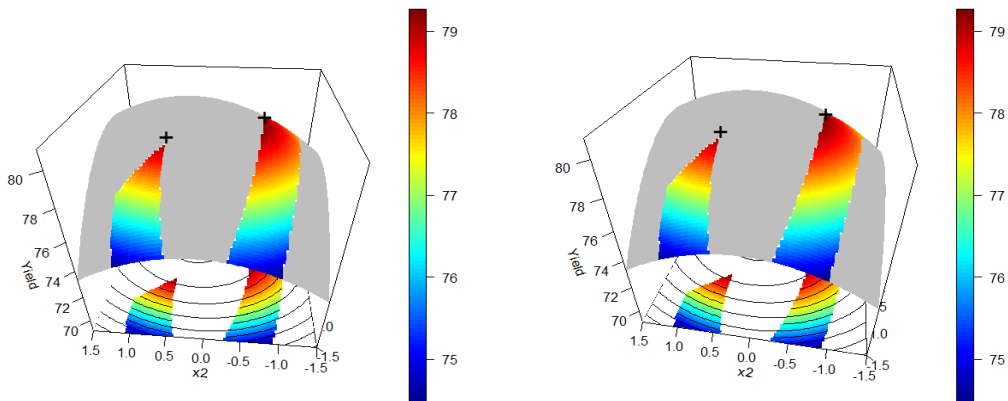


Figure 2.11. The stereo image for region of the optimum found by yield response surfaces.

알 수는 있지만 3차원 그림을 2차원으로 완벽하게 표현할 수 없기 때문에 그림에 왜곡이 생길 수밖에 없어서 그림을 보더라도 최적해의 위치를 명확하게 알기는 어렵다. 하지만 VR을 이용하여 살펴볼 경우 이러한 왜곡 없이 3차원 그림을 온전히 3차원으로 볼 수 있기 때문에 두 최적해가 어디에 위치하고 있는지 직관적으로 파악할 수 있다.

2.2. Real-time method

Real-time method는 실시간으로 동적 이미지를 VR로 확인하는 방법이다. 공유 방법에는 Remote control method와 Dual monitor method가 있다. Remote control method는 스마트폰을 컴퓨터 제어 기기로 사용하는 방식이다. 반면 Dual monitor method는 컴퓨터에서 실행한 R 작업을 컴퓨터 모니터 화면뿐만 아니라 스마트폰 화면에서도 작업 결과를 동시에 보는 방식이다. 두 방식 모두 컴퓨터 모

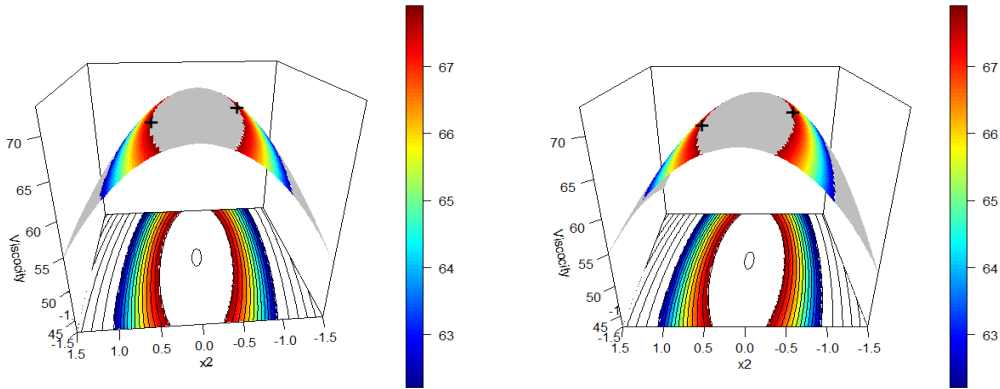


Figure 2.12. The stereo image for region of the optimum found by viscosity response surfaces.

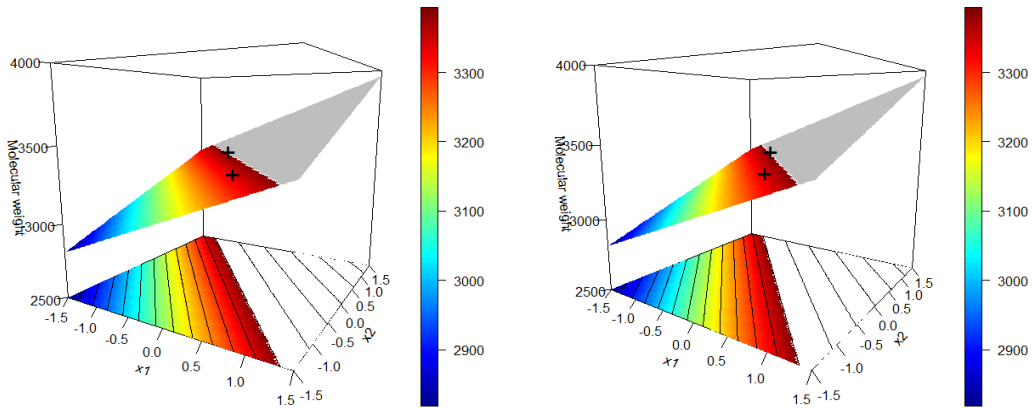


Figure 2.13. The stereo image for region of the optimum found by molecular weight response surfaces.

니터 화면과 스마트폰 화면을 동시에 공유하므로 VR 기기를 이용하면 컴퓨터 모니터에서 생성되는 동적 이미지가 스마트폰 화면에 실시간 입체적으로 나타난다. Real-time method를 이용하여 동적 이미지를 살펴볼 때 두 가지 방법이 있다. 첫 번째는 데이터로부터 얻은 3차원 동적 이미지를 360도로 회전시켜서 보는 것이다. 두 번째 방법은 슬라이드 바를 이용하여 원하는 각도의 동적 이미지를 살펴보는 것이다. 이 때 슬라이드 바를 사용하게 되면 그림을 좌·우 또는 위·아래로 회전하거나 크기를 조절하는 등 원하는 값들을 조절할 수 있도록 설정할 수 있다. Real-time method에 대한 구체적인 절차는 아래의 예제를 통하여 소개하고자 한다.

다섯 번째 예제는 다음과 같이 각각 모평균벡터와 모공분산행렬을 갖는

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

두 개의 이변량 정규 분포들에 대하여 동적 이미지인 Figures 2.15–2.18을 그릴 것이다. Real-time method를 이용하여 수행하는 과정은 다음과 같으며 Figure 2.14를 통하여 구체적인 절차를 확인할 수

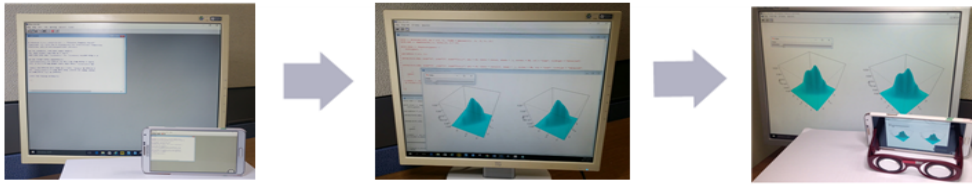


Figure 2.14. The procedure of Real-time method for the bivariate normal random variates $BN(0,0,1,1,0.9)$.

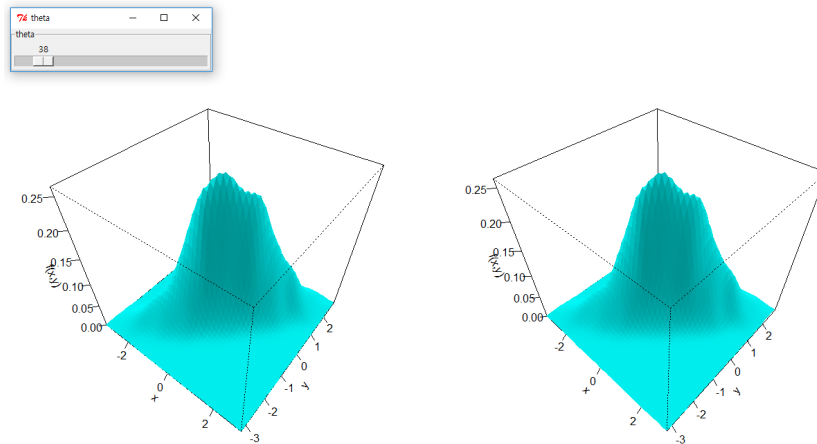


Figure 2.15. The stereo image for the bivariate normal random variates $BN(0, 0, 1, 1, 0.9)$ (1).

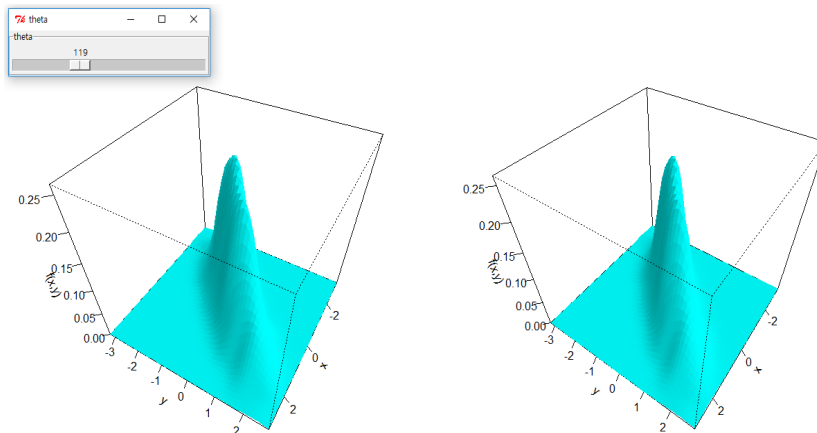


Figure 2.16. The stereo image for the bivariate normal random variates $BN(0, 0, 1, 1, 0.9)$ (2).

있다. R 라이브러리는 MASS, rpanel, aplpack를 이용하였다.

단계 1: 컴퓨터 모니터 화면과 스마트폰 화면을 실시간으로 공유한다.

단계 2: R 프로그램을 이용하여 이변량 정규 분포의 데이터를 분석하고 동적 이미지를 생성한다.

단계 3: 스마트폰을 VR 안경 또는 헤드셋에 끼워둔 채로 슬라이드 바를 움직여서 다양한 각도의 이변

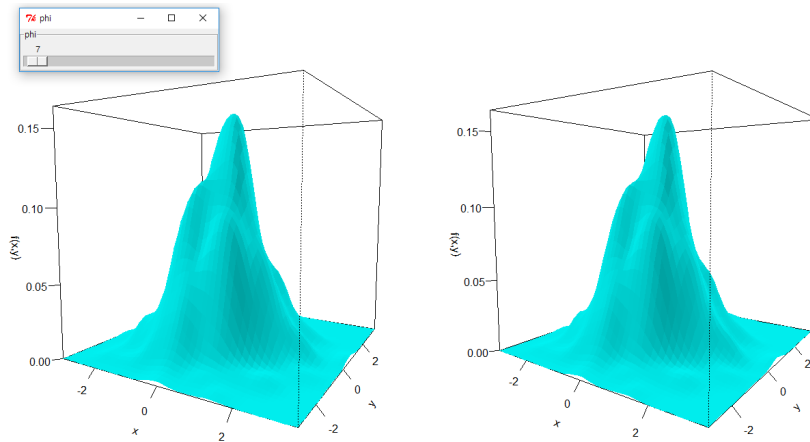


Figure 2.17. The stereo image for the bivariate normal random variates $BN(0, 0, 1, 1, 0)$ (1).

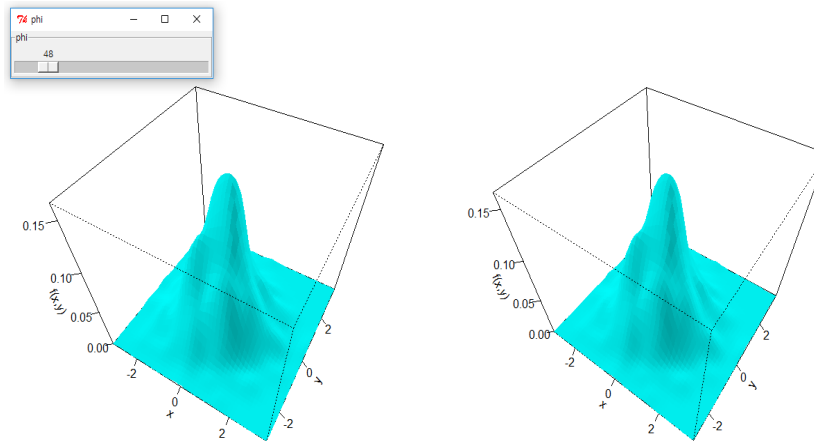


Figure 2.18. The stereo image for the bivariate normal random variates $BN(0, 0, 1, 1, 0)$ (2).

량 정규 분포 동적 이미지를 입체적으로 확인한다.

Figures 2.15와 2.16은 슬라이드 바를 이용하여 이변량 정규분포 $BN(0, 0, 1, 1, 0.9)$ 를 좌·우로 회전시켜서 살펴본 동적 이미지이고 Figures 2.17과 2.18은 슬라이드 바를 이용하여 이변량 정규분포 $BN(0, 0, 1, 1, 0)$ 를 위·아래로 회전시켜서 살펴본 동적 이미지이다. 회전시킬 때뿐만 아니라 정지해 있는 경우에도 그림을 입체로 볼 수 있다. 컴퓨터 모니터 화면과 스마트폰 화면을 실시간으로 공유한 채로 VR 기기와 슬라이드 바를 이용하여 원하는 각도의 그림을 조절하면서 볼 경우 Figures 2.15-2.18의 3차원 구조를 좀 더 자세히 살펴볼 수 있다.

여섯 번째 예제는 다음과 같이 각각 모평균벡터와 모공분산행렬을 갖는 두 개의 이변량 정규분포

$$\mu_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_3 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}, \quad \mu_4 = \begin{pmatrix} 3 \\ 3 \end{pmatrix}, \Sigma_4 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

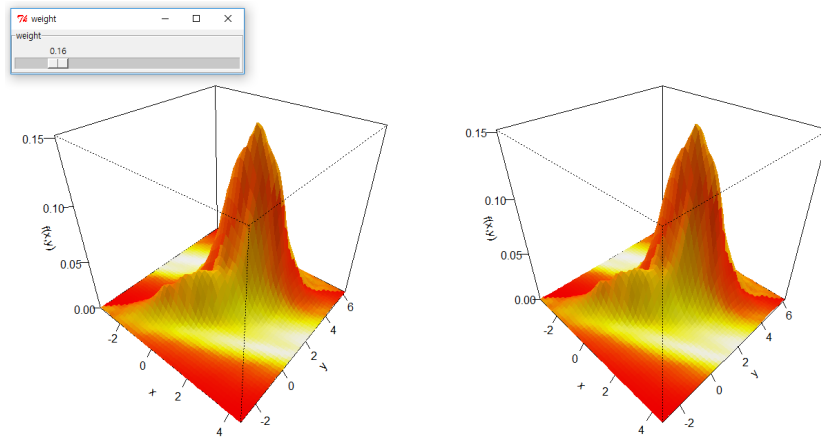


Figure 2.19. The stereo image for the mixture of the bivariate normal random variates $BN(0, 0, 1, 1, 0.9)$ and $BN(3, 3, 1, 1, 0.5)$ with 0.16 weight (1).

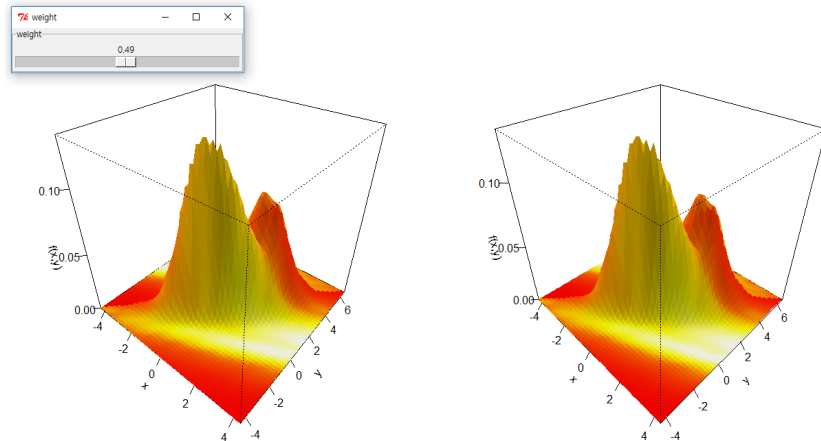


Figure 2.20. The stereo image for the mixture of the bivariate normal random variates $BN(0, 0, 1, 1, 0.9)$ and $BN(3, 3, 1, 1, 0.5)$ with 0.49 weight (2).

에 대하여 w 를 다르게 둔 이변량 혼합 정규분포의 동적 이미지인 Figures 2.19–2.20을 그린 것이다. 이변량 혼합 정규분포에 대한 확률밀도함수는 아래와 같이 주어진다.

$$f(\mathbf{x}; w, \mu_3, \mu_4, \Sigma_3, \Sigma_4) = w \frac{1}{(2\pi)^2 |\Sigma_3|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \mu_3)\Sigma_3^{-1}(\mathbf{x} - \mu_3)} + (1 - w) \frac{1}{(2\pi)^2 |\Sigma_4|^{1/2}} e^{-\frac{1}{2}(\mathbf{x} - \mu_4)\Sigma_4^{-1}(\mathbf{x} - \mu_4)}, \quad (2.3)$$

여기서 w 는 weight이다. w 를 조절할 수 있도록 설정한 슬라이드 바를 이용하여 다양한 이변량 혼합 정규분포 그림을 그릴 수 있다. MASS, rpanel, aplpack R 패키지를 이용했으며 생성된 3차원 동적 이미지는 VR을 통하여 온전히 3차원으로 살펴볼 수 있다. 따라서 3차원 동적 이미지를 한 눈에 살펴볼 수 있기 때문에 이변량 혼합 정규분포에 대한 정보를 빠르게 이해할 수 있다. 이 작업을 위한 R 코드는 부록 A.3에 주어져 있다.

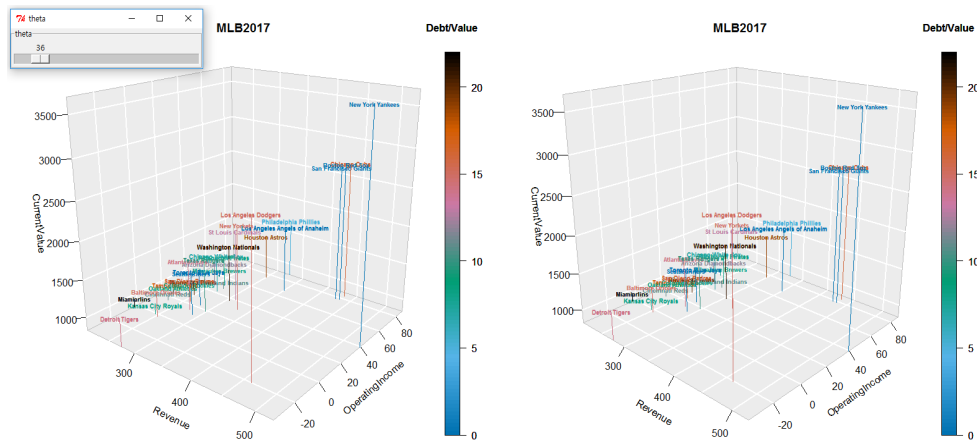


Figure 2.21. The stereo image for MLB 2014 (1).

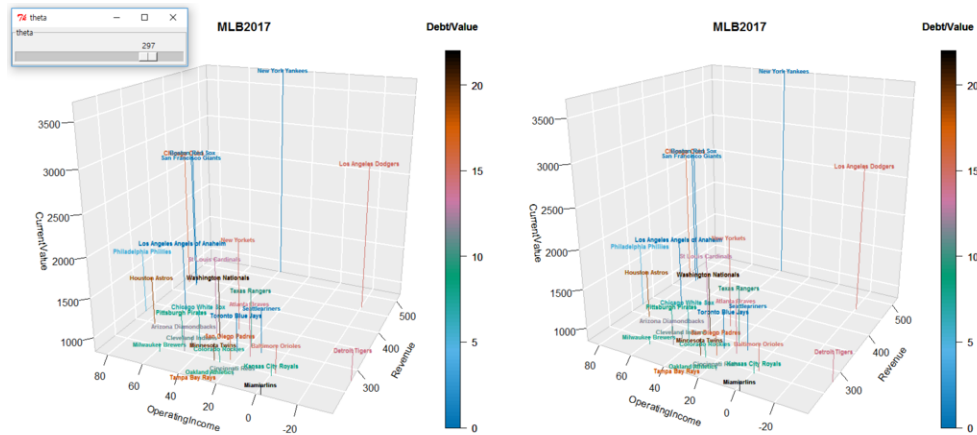


Figure 2.22. The stereo image for MLB 2014 (2).

일곱 번째 예제는 Forbes지(www.forbes.com)에서 제공한 2014년 미국 메이저리그 프로야구팀에 대한 경영평가 자료를 이용하여 분석한 것이다. 이 데이터는 30팀에 대하여 7개의 변수(순위(rank based on current value), 팀 이름(team name), 현재 구단 가치(current value, \$ mil), 1년 구단 가치 변화율(one-year value change, %), 가치 대비 부채비율(debt/value, %), 수입(revenue, \$ mil), 운영 수익(operating income, \$ mil))로 구성되어 있다.

Figures 2.21과 2.22는 메이저리그의 경영실태평가 자료 중 3개의 변수(현재 구단 가치, 수입, 운영 수익)에 대하여 그린 3차원 산점도이며 좌·우를 회전한 그림이다. 컬러범례는 네 번째 변수인 가치 대비 부채비율(debt/value (%))이다. Real-time method를 이용하여 3차원 산점도의 입체 구조를 확인할 수 있으며 좌·우를 조절할 수 있도록 설정한 슬라이드 바를 이용하여 원하는 각도의 그림을 확인할 수 있다. 따라서 VR을 통하여 원하는 자료가 어디에 분포하였는지 입체감을 유지한 채 동적으로 자세히 살펴볼 수 있다. 이 작업을 위한 R 코드는 부록 A.4에 주어져 있다.

3. 결론

3차원 그림을 2차원으로 표현할 경우 완벽하게 그림을 표현할 수 없기 때문에 그림에 왜곡이 생겨서 부차적인 확인을 통하여 그림을 확인해야 하는 번거로움이 생기게 된다. 이에 대하여 본 논문에서는 2차원 평면인 컴퓨터 모니터 화면보다 VR을 이용하여 입체적으로 살펴본 3차원 그래프가 좀 더 정확하게 정보를 제공 할 수 있다는 것을 살펴보았다. 그리고 이를 구현하는 두 가지 방법인 Batch method 와 Real-time method에 대하여 논의하였다. VR를 이용할 경우 아직까지 어렵고 콘텐츠가 제한적이라는 한계가 있지만 그 가능성은 무시할 수 없다. 몇 년 전까지만 해도 VR 기기가 비싸고 무겁고 거추장스러운 단점이 있었지만 현재는 많이 저렴해지고 가벼워졌으며 고성능 렌즈로 만들어졌기 때문에 선명하게 그림을 분석할 수 있게 되었다. 그리고 거의 모든 사람들이 스마트폰을 사용하고 있기 때문에 어려움 없이 VR 기기를 이용하여 3차원 그림을 분석할 수 있다. 2016년에 발표한 가트너(Gartner)의 하이퍼 사이클(Hype Cycle) 사이트(<http://www.gartner.com>)에 의하면 현재 가상현실은 각성의 단계('Slope of Enlightenment')에 있으며 기술이 안정기로 접어들어 상용화되는 시점은 향후 5-10년 사이로 예상하고 있다고 발표하였다. 이처럼 현재 삼성, 구글, 페이스북 등 여러 IT 선진 기업들이 앞다투어 가상현실 시대를 공격적으로 대비하고 있다. 따라서 VR 기기를 이용할 때 겪는 한계는 곧 극복할 수 있을 것이다. 또한 3차원 그림을 전체적인 조명도로만 보는 것이 아니라 직접 움직이면서 3차원 그림을 온전히 3차원으로 볼 때 자료의 분석이 훨씬 효과적이라는 것은 자명한 사실이다. 아직까지 VR이 게임 및 영상 분야 등에 한정되어 많은 발전이 있지만 이러한 콘텐츠들의 발전 방향은 곧 교육, 산업 및 다양한 연구 분야로 이어질 것이며 통계적 시각화에도 긍정적인 효과를 가져다줄 것으로 예상된다. 현재 사용자가 3차원 물체를 만지고 조작하며 별도의 입력기기 없이 가상환경과 상호작용이 가능한 동적 기술과 3차원 가상 환경에서 공동 작업이 가능하도록 하는 다중 사용자 환경 기술 등 새로운 형태의 기술들이 개발되고 있다. 이러한 서비스를 이용하여 원하는 자료를 3차원 통계적 시각화를 통하여 구체적으로 자료를 살펴볼 수 있을 것이며 여러 사람들이 3차원 가상 환경에서 공동으로 3차원 그림을 작업하고 공유하는 등 효과적인 자료 분석이 가능할 것이다. 따라서 VR은 3차원 데이터 시각화 분석기술의 발전에 디딤돌이 될 수 있을 것이다.

부록 A: R 코드

A.1. 예제들을 구현하기 위한 R 라이브러리

```
install.packages('MASS'); library(MASS)
```

```
# 삼차원 그림을 그리기 위한 라이브러리
```

```
install.packages('rgl'); library(rgl)
```

```
# 삼차원 그림을 그리기 위한 라이브러리 (고정형)
```

```
install.packages('plot3D'); library(plot3D)
```

```
install.packages('scatterplot3d'); library(scatterplot3d)
```

```
# 슬라이드바 생성 라이브러리 (타입 1)
```

```
install.packages('rpanel'); library(rpanel)
```

```
# 슬라이드바 생성 라이브러리 (타입 2)
install.packages('aplpack'); library(aplpack)
```

A.2. Figure 2.3을 구현하기 위한 R 코드

```
attach(iris)
v1 = Sepal.Length ; v2 = Sepal.Width ; v3 = Petal.Length ; v4 = Petal.Width
lab_v1 = "꽃받침 길이"; lab_v2 = "꽃받침 폭"; lab_v3 = "꽃잎 길이"; lab_v4 = "꽃잎 폭"

par(mfrow=c(1,2))
scatter3D(v1, v3, v4, phi = 20, theta = 40, pch= 19, cex = 1, bty = "g", xlab = lab_v1, ylab =
lab_v3, zlab = lab_v4, ticktype = "detailed", col = c(rep("red", 50),rep("green", 50),rep("blue",
50)), colkey = F, cex.lab = 1)
legend(-0.1, -0.2, col = c("red", "green", "blue"), pch = c(19,19,19), c("setosa", "versicolor",
"virginica"), cex= 1,y.intersp = 1)
scatter3D(v1, v3, v4, phi = 20, theta = 45, pch=19, cex = 1, bty = "g", xlab = lab_v1, ylab =
lab_v3, zlab = lab_v4, ticktype = "detailed", col = c(rep("red", 50),rep("green", 50),rep("blue",
50)), colkey = F, cex.lab = 1)
legend(-0.1, -0.2, col = c("red", "green", "blue"), pch = c(19, 19, 19), c("setosa", "versicolor",
"virginica"), cex=1,y.intersp = 1)
```

A.3. Figure 2.19와 Figure 2.20을 구현하기 위한 R 코드

```
loc.mix <- function(n, p, mu1, mu2, Sigma1, Sigma2) {
# generate the normal mixture data
n1 <- rbinom(1, size=n, prob=p)
n2 <- n-n1
x1 <- mvrnorm(n1, mu=mu1, Sigma1)
x2 <- mvrnorm(n2, mu=mu2, Sigma2)
X <- rbind(x1, x2) # combine the samples
return(X[sample(1:n),]) # mix them
}
bivn3 <- mvrnorm(10000, mu = c(0, 0), Sigma = matrix(c(1, .9, .9, 1), 2))
bivn4 <- mvrnorm(10000, mu = c(1, 3), Sigma = matrix(c(1, .5, .5, 1), 2))
bivn_mixture <- rbind(bivn3, bivn4)
bivn_mixture.kde <- kde2d(bivn_mixture[,1], bivn_mixture[,2], n = 50)

# generate the normal mixture data
```

```

n <- 10000; k=0.3; p <- c(k, 1-k); mu <- matrix(c(0,0,1,3),2,2)
Sigma1 <- matrix(c(1, .9, .9, 1), 2); Sigma2 <- matrix(c(1, .5, .5, 1), 2)
i <- sample(1:2, replace=TRUE, prob=p, size=n)
k <- table(i)

x1 <- mvrnorm(k[1], mu=mu[1,], Sigma1)
x2 <- mvrnorm(k[2], mu=mu[2,], Sigma2)
bivn_mixture.ex <- rbind(x1, x2) # the mixture data
bivn_mixture.kde.ex <- kde2d(bivn_mixture.ex[,1], bivn_mixture.ex[,2], n = 50)
col1 <- rainbow(length(bivn_mixture.kde.exz))[rank(bivn_mixture.kde.exz)]
col2 <- heat.colors(length(bivn_mixture.kde.exz))[rank(bivn_mixture.kde.exz)]

plot.draw <- function(panel)
with(panel, {
par(mfrow = c(1, 2))
bivn_mixture.ex <- (loc.mix(10000, weight, c(0,0), c(1, 3), Sigma1=matrix(c(1, .9, .9, 1), 2),
Sigma2=matrix(c(1, .5, .5, 1), 2)))
bivn_mixture.kde.ex <- kde2d(bivn_mixture.ex[,1], bivn_mixture.ex[,2], n = 50)

persp(bivn_mixture.kde.ex, xlab="x", ylab="y", zlab="f(x,y)", phi = 35, theta = 40, shade = .1,
border = NA, col = col2, ticktype = "detailed" )
persp(bivn_mixture.kde.ex, xlab="x", ylab="y", zlab="f(x,y)", phi = 35, theta = 45, shade = .1,
border = NA, col = col2, ticktype = "detailed")
})
panel
}
a.panel <- rp.control("weight")
rp.slider(a.panel, weight, from=0, to=1, title="weight", .action=plot.draw, showvalue=TRUE)

```

A.4. Figure 2.21과 Figure 2.22를 구현하기 위한 R 코드

```

MLB2014=read.csv("c:/WORK/MLB2014.csv", header=T, sep=",")
attach(MLB2014)
plot.draw <- function(panel) {
with(panel, {
par(mfrow = c(1, 2))
with(MLB2014, text3D(Revenue, OperatingIncome, CurrentValue, colvar = DebtoverValue, col =
gg.col(100), theta = theta1, phi = 20, xlab = "Revenue", ylab = "OperatingIncome", zlab =

```



```

“CurrentValue”, main = “MLB2014”, labels = Team, cex = 0.6, bty = “g”, ticktype = “detailed”,
d = 2, clab = c(“Debt/Value”), adj = 0.5, font = 2))
with(MLB2014, scatter3D(Revenue, OperatingIncome, CurrentValue - 1, colvar = DebtoverValue,
col = gg.col(100), type = “h”, pch = “.”, add = TRUE))
with(MLB2014, text3D(Revenue, OperatingIncome, CurrentValue,colvar = DebtoverValue, col =
gg.col(100), theta = theta1+5, phi = 20,xlab = “Revenue”, ylab = “OperatingIncome”, zlab =
“CurrentValue”, main = “MLB2014”,labels = Team, cex = 0.6,bty = “g”, ticktype = “detailed”,
d = 2,clab = c(“Debt/Value”), adj = 0.5, font = 2))
with(MLB2014, scatter3D(Revenue, OperatingIncome, CurrentValue - 1,colvar = DebtoverValue,
col = gg.col(100),type = “h”, pch = “.”, add = TRUE))
})
panel
}
a.panel <- rp.control(“theta”)
rp.slider(a.panel, theta1, 0,360, title=“theta”,action=plot.draw, showvalue=TRUE)

```

References

- Bowman, A. (2015). rpanel: Simple interactive controls for R using the tcltk library. R package version 1.1-3.
- Campos, M. M. (2007). Way Cooler: PCA and Visualization Linear Algebra in the Oracle Database 2, <http://oracledmt.blogspot.kr/2007/06/way-cooler-pca-and-visualization-linear.html>.
- Ligges, U. (2017). scatterplot3d: 3D Scatter Plot. R package version 0.3-38.
- Murdoch, D. (2017). rgl: 3D Visualization Using OpenGL. R package version 0.97.0.
- Myers, R. H., Montgomery, D. C. and Anderson-Cook, C. M. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd ed, Wiley, New York.
- Ripley, B. (2016). MASS: Support Functions and Datasets for Venables and Ripley’s MASS. R package version 7.3-47.
- Sarkar, D. (2016). lattice: Trellis Graphics for R. R package version 0.20-35.
- Soetaert, K. (2016). plot3D: Plotting Multi-Dimensional Data. R package version 1.1.
- Wolf, H. P. (2015). aplpack: Another Plot PACKage: stem.leaf, bagplot, faces, spin3R, plotsummary, plothulls, and some slider functions. R package version 1.3.0.
- http://astrostatistics.psu.edu/datasets/SDSS_quasar.html.
- http://forbes.com/mlb_valuations/list.
- <http://gartner.com/newsroom/id/3412017>.

가상현실을 이용한 3차원 데이터 시각화

이재은^a · 안소진^a · 장대흥^{a,1}

^a부경대학교 통계학과

(2017년 02월 13일 접수, 2017년 04월 12일 수정, 2017년 05월 04일 채택)

요약

자료 분석가가 대용량 데이터를 분석할 때 이해하기 쉽고 효과적으로 전달하기 위하여 다양한 데이터 시각화 방법들을 사용하고 있다. 그 중 3차원 그림은 2차원 화면상에서 입체로 표현하기 위하여 그림을 회전시켜야 한다. 이러한 단점을 보완하기 위하여 본 연구에서는 3차원 시각화 자료를 입체적으로 살펴보는 방법으로 가상현실을 이용한 두 가지 방식인 Batch method와 Real-time method를 제안하였다. 이 방법들을 통하여 3차원 시각화 자료를 입체적으로 살펴볼 때 3차원 구조를 좀 더 명확하게 확인할 수 있을 것이다.

주요용어: 데이터 시각화, 가상현실, 스테레오 이미지, R 패키지

이 논문은 부경대학교 자율창의학술연구비(2016년)에 의하여 연구되었음.

¹교신저자: (48513) 부산광역시 남구 용소로 45, 부경대학교 통계학과. E-mail: dhjang@pknu.ac.kr