



Interactive Data Visualization Based Realtime Monitoring and Fault Detection System

Seung-Taek Ryo^{*}

School of Computer Engineering, HanShin University

A B S T R A C T

Effective representation of various types of data from various types of equipment is not a trivial task. The more data items collected for sophisticated data analysis, the more trouble the system operator has to identify important items. In order to solve these problems, it is necessary to develop a real-time interactive data visualization system that can be efficiently observing large-volume data at a glance and confirming fault detection in advance. In this paper, we propose a real - time observation and fault detection system using interactive data visualization method to efficiently manage consecutively generated raw data. The proposed system consists of data collection, data visualization, data fault detection and alarm steps. In the data collection step, the real-time generated device data using the data collection device is stored, and the collected data is massaged and generalized through data preprocessing process. The normalized data is subjected to data analysis, visualization mapping, and rendering in the data visualization step. To visualize and render the defined data primitives, we use heat map, radar chart, and stream graph. In addition, statistical process control (SPC) rule-based fault detection and alarm methods are used to detect data errors. The proposed method can be widely used in various fields such as image generation, data fusion, fault detection and machine learning.

© 2018 KKITS All rights reserved

KEYWORDS : Realtime rendering, Data visualization, Temporal analysis, Spatial analysis, Data fusion

ARTICLE INFO: Received 5 June 2018, Revised 28 June 2018, Accepted 10 August 2018.

^{*}Corresponding author is with the School of Computer Engineering, HanShin University, 137 Hanshin Univ.-gil

Osan-si, Gyeonggi-do, 18101, KOREA.
E-mail address: stryoo@hs.ac.kr

1. 서론

대 다수의 산업현장은 다양한 설비로부터 여러 형태의 데이터가 빈번하게 발생하고 있다[1]. 이러한 대용량 데이터를 실시간으로 분석하여 오류감지 및 설비고장을 미리 관측하고 예견할 수 있는 검출 시스템 개발에 대한 연구가 활발히 진행되고 있다[2]. 정교한 데이터 분석 등을 위해 수집하는 데이터 항목이 많아질수록 시스템 운영자가 중요한 항목을 식별하는데 장애요인으로 작용한다. 이러한 문제점을 개선하기 위해서는 대용량 데이터를 효과적으로 한 눈에 관측하고 미리 오류 감지를 확인할 수 있는 실시간 인터랙티브 데이터 시각화 기술 개발이 필요하다[3, 4]. 본 논문에서는 대용량 데이터 효율적 관리를 위해 데이터 시각화 기술을 이용한 실시간 관측 및 오류 감지 시스템을 개발하고자 한다.

2. 관련연구

데이터 시각화 방법의 유형은 구조적 데이터(지질, 의료 등)를 이용한 과학적 시각화(Scientific Visualization), 비구조적 데이터(뉴스, 주식, 페이스북 등)를 이용한 정보 시각화(Information Visualization), 멀티모달 데이터(텍스트, 이미지, 오디오, 비디오 등)를 이용한 시각적 분석(Visual Analytics)으로 나눌 수 있다.

과학적 시각화[5]는 물리적 공간, 즉 시공간 데이터 (2차원~4차원)에서 잘 정의 된 데이터를 다룬다. 이러한 데이터는 연속 공간에서 연속적인 이벤트로 발생하나 데이터 시각화를 위해 연속적 데이터를 샘플링하여 이산적으로 표현한다. 이러한 방법은 물리학, 화학, 지리학, 의학, 생물학, 기후, 해양학, 에너지 등과 같은 분야에서 측정된 데이터를 시뮬레이션 하는데 사용되고 있다.

정보 시각화[6]는 고차원(4차원이상), 추상적 데이터(트리, 그래프, 계층 구조 등)를 다루며 데이터 자체가 이산적으로 발생한다. 과학적 시각화와 달리 데이터 분석을 위해 특징들을 정의하기가 어렵다. 금융, 마케팅, 통계, 소셜 미디어, 정치 등과 같은 분야에서 사용되고 있다.

시각적 분석[7]은 정보 시각화 및 과학 시각화 분야에서 파생된 분야로 분석 추론에 중점을 두고 있다. 이 방법은 오디오 클립, 비디오, 사진과 같은 멀티 모달 데이터를 다룬다. 분석 기술과 대화형 시각화 기술을 통합하여 과학, 비즈니스 및 분석 분야에서 점차 증가하는 질문에 대한 해결 방법을 찾기 위해 사용되고 있다.

본 논문에서는 데이터 수집 장치로부터 시간적 대용량 데이터를 획득하여 실시간 데이터 가시화 기술을 통한 시각적 분석 기반 실시간 관측 및 오류 감지 시스템을 제안한다.

3. 대화형 데이터 시각화 기반 실시간 관측 및 오류 감지 시스템

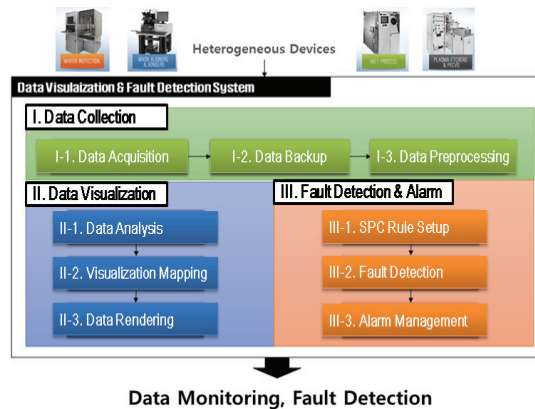


그림 1. 대화형 데이터 시각화 기반 실시간 관측 및 오류 감지 시스템의 구성도

Figure 1. Flowchart of interactive data visualization based realtime monitoring and fault detection system

실시간 관측 및 오류 감지 시스템은 수십 개 이상의 항목을 한눈에 관측할 수 있도록 표준화/정규화를 통한 실시간 데이터 통합 시각화가 가능해야 하며, 2차원 도표 형태(꺾은선, 막대, 파이 차트 등) 표현이 아닌 데이터의 중요도에 따른 시각화 기술을 사용하여야 한다.

또한, 정적 데이터의 고정된 나열이 아닌 시간/일/월/연도별 상호작용이 필요하며 개별 데이터로부터 획득된 2차원 시각화 영상을 3차원 통합 볼륨으로 생성하여 원하는 시점의 단면(Cross-Section)을 분석할 수 있어야 한다. 마지막으로, 오류 감지 알고리즘을 통해 설비데이터의 항목별 이상 여부 및 설비 고장을 실시간으로 관찰 및 감지가 가능하여야 한다.

인터랙티브 데이터 시각화 기술을 이용한 대용량 설비 데이터의 실시간 관측 및 오류 감지 시스템은 <그림 1>과 같이 데이터 수집, 데이터 시각화 와 데이터 오류 감지 및 알람 단계로 나뉜다.

데이터 수집 단계는 데이터 획득, 데이터 저장, 데이터 전처리 과정으로 구성된다. 데이터 획득 및 저장 과정에서 데이터 수집 장치를 이용하여 발생한 실시간 설비 데이터를 저장한다. 이렇게 획득된 데이터는 전처리 과정에서 데이터 메시지와 정규화(일반화) 과정을 수행한다. 데이터 시각화 단계는 데이터 분석, 시각화 매핑, 데이터 렌더링 과정을 거친다. 데이터 분석 과정에서는 인자별 상관 및 회귀분석을 통한 주요 인자를 추출하고 주성분 분석 기술을 통해 데이터 차원을 축소한다. 시각화 매핑 과정에서 데이터 차원 및 타입에 따라 시각화 프리미티브를 정의하고 데이터 렌더링 과정에서 매핑된 프리미티브를 이용하여 2차원 영상으로 데이터를 시각적으로 표현한다. 마지막 데이터 오류 감지 및 알람 단계는 공정제어 정의, 오류 감지, 알람 관리 과정으로 구성된다. 공정제어 정의 과정에서는 설비데이터 항목별 규칙을 정의한다. 오류

감지 및 알람 과정에서는 정의된 규칙을 바탕으로 항목별 오류 감지에 대한 알람 처리한다.

3.1 대용량 설비 데이터 수집

대다수의 산업현장은 다양한 설비로부터 여러 형태의 데이터가 빈번하게 발생하고 있다. 이러한 다양한 데이터를 활용한 솔루션 요구가 증대되고 있으나 표준화되어 있지 않은 데이터 수집 시스템과 수많은 형태의 데이터 수집 장치들로 인해 데이터 수집과 관리가 어려운 현실이다.

설비 별로 다양한 데이터 통신 프로토콜 사용하고 있어 대용량 설비 데이터를 효율적으로 수집하기 위해 모델링 서버 및 수집 장치 개발이 필요하다. 대용량 설비 데이터를 수집하기 위한 환경은 <그림 2>와 같이 모델링 서버, 데이터 수집장치, 대상 설비로 구성된다.

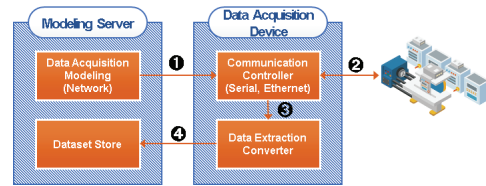


그림 2. 설비로부터 데이터 세트의 수집
Figure 2. Collection of data sets from the equipment

데이터 수집 모델링 과정에서 대상 생산설비에 따른 통신 및 데이터 수집 모델 정의하고, 통신 제어 과정에서 데이터 수집 모델에 따라 설비와 데이터 통신(시리얼, 이더넷 등)을 수행한다. 데이터 추출/변환 과정에서는 생산설비로부터 수신된 메시지를 분석하여, 데이터 추출하고 용도에 맞게 변환한다. 데이터 저장 과정에서는 생산설비로부터 데이터 수집 장치에서 추출/변환된 데이터를 데이터 베이스에 저장한다.

수집된 데이터를 분석하기 위해서는 데이터 전

처리 과정이 필요하다. 데이터 전처리 단계는 데이터 마사징, 데이터 샘플링/필터링, 데이터 표준화(Data Standardization), 데이터 정규화(Data Normalization)로 구성된다. 데이터 마사징 과정에서 수집된 데이터에서 의미가 있는 데이터 구간을 선별하고 데이터 샘플링/필터링 과정에서 데이터의 일부를 추출하여 이상치(Outlier)를 제거하거나 증폭한다. 데이터 표준화 과정에서 데이터의 분포를 평균값/중간값, 표준편차 등을 이용하여 유사하게 만들고 데이터 정규화 과정에서 데이터의 범위를 인자별 최소/최대값을 이용하여 일치시킨다.

3.2 인터랙티브 데이터 시각화

대용량 설비 데이터를 시각화하기 위해서는 데이터 분석을 통한 주요인자 추출 과정이 필요하다. 주요 인자를 추출하기 위해 두 인자간의 연관된 정도인 상관계수(Correlation Coefficient)를 통해 인자별 상관분석 (Correlation Analysis)을 수행하거나 두 인자간의 원인과 결과의 인과관계(회귀식, 회귀계수 추정 등)를 통해 인자별 회귀분석 (Regression Analysis) 수행한다[8]. 이러한 통계적 분석(주요인자별 관계에 따라 데이터 정렬) 과정을 통해 주요 인자를 선정할 수 있다.

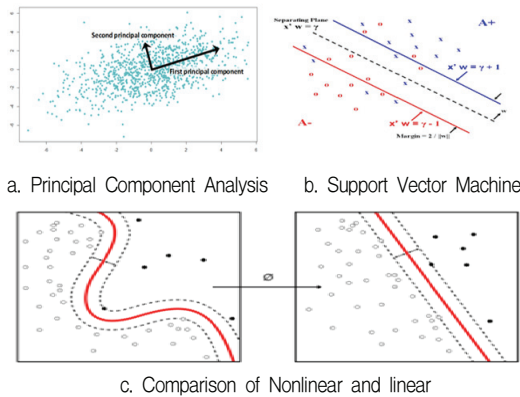


그림 3. 데이터 차원의 축소
Figure 3. Reduction of Data Dimension

다차원의 주요 인자들 분석하기 위해서는 데이터 차원 축소가 필요하다. 데이터 차원 축소를 위해 서로 상관성이 높은 여러 변수들의 선형조합으로 만든 새로운 변수들로 축소하는 기법인 주성분 분석(PCA: Principal Component Analysis)[9]을 사용하거나 변수들 간의 상관관계를 고려하여 서로 유사한 변수들 끼리 묶어주는 기법인 요인 분석 (Factor Analysis)을 사용한다.

시각화 모델링 과정에서는 데이터 차원 및 타입에 따른 분석을 통해 시각화 프리미티브를 정의한다. 추출된 주요인자를 효과적으로 표현하기 위해 데이터의 특성에 맞게 기하정보, 색상, 채도, 밝기, 크기, 위치 등과 같은 프리미티브 정보를 정의하여 시각화 매핑(Visualization Mapping)을 수행한다.

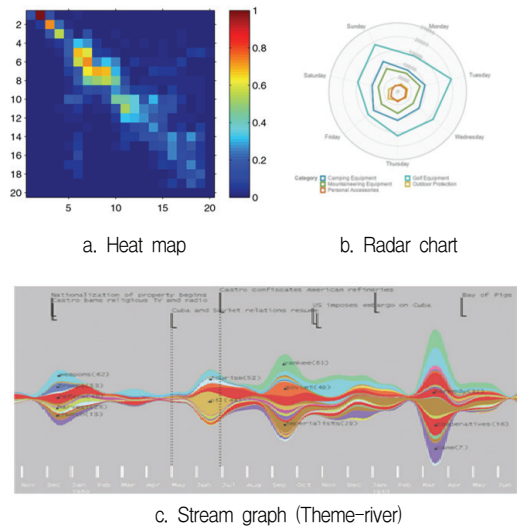


그림 4. 데이터 시각화 (HeatMap, Radar chart, stream graph)
Figure 4. Data Visualization (HeatMap, Radar chart, stream graph)

본 논문에서는 정의된 데이터 프리미티브를 시각화 렌더링하기 위한 방법으로 매트릭스에 포함된 개별 데이터의 수치값을 중요도에 따라 색상으로 매핑하여 표현하는 시각화 방법인 히트맵(Heat

Map)[10], 시계열 다중인자를 표현하기 위한 2차원 방사형 모델로 표현하는 시각화 방법인 레이더 차트(Radar Chart)[11], 시간의 흐름에 따라 변화하는 항목을 흐르는 강물처럼 표현하는 시각화 방법인 스트림 그래프(Stream Graph)[12] 방법을 사용하였다(그림 4).

실시간 관측 및 상호작용이 가능하도록 하기 위해서는 인터랙티브 데이터 시각화 기술이 필요하다. 본 논문에서는 이를 위해 정적 데이터의 고정된 나열이 아닌 시간/일/월/년도별 상호작용(Interactive) 캘린더 뷰 기술을 사용하고 개별 데이터 시각화 영상의 3D 통합볼륨 생성을 통해 원하는 시점의 단면(Cross-Section) 분석 기술 사용하였다.

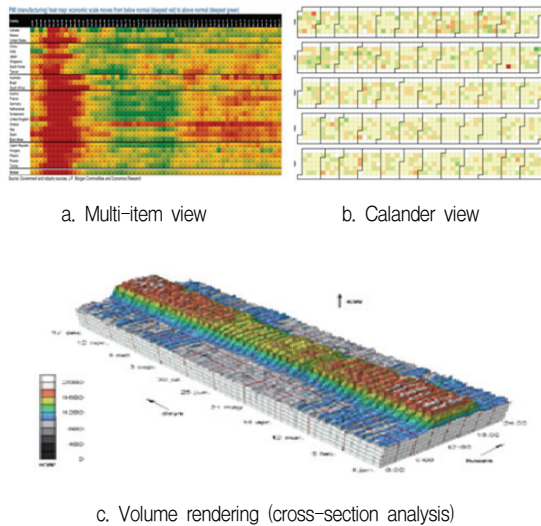


그림 5. 실시간 인터랙티브 데이터 시각화
Figure 5. Realtime Interactive Data Visualization

3.3 데이터 오류 감지 및 알람

데이터의 오류 감지를 하기 위해 통계적 공정 관리(SPC: Statistical Process Control)[13] 규칙 기반 오류 감지 및 알람 시스템을 개발하였다.

공정에 대한 설비데이터 항목별 SPC 규칙을 정의하고 다양하고 복잡한 상황에 대한 규칙 정의를 보다 편리하게 정의하기 위해 단위 규칙들의 조합을 통해 쉽게 생성하고 시뮬레이션할 수 있게 하였다. 이를 위해 간단한 SPC 규칙을 조합하여 복잡한 규칙을 정의할 수 있는 SPC 규칙트리(RuleTree) 기술을 사용하였다. 정의된 SPC RuleTree를 바탕으로 항목별 오류 미리 감지할 수 있으며 감지된 오류에 대한 다양한 알람 처리(DB저장, 문자 발송 시스템과 통합, 경고음 울림 등) 구현하였다.

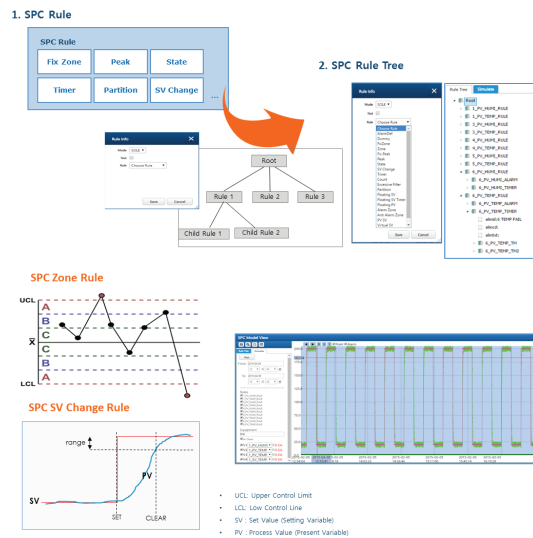


그림 6. 통계적 공정제어 규칙 트리 설정 및 시뮬레이션
Figure 6. SPC Rule Tree Setup & Simulation

4. 구현 결과

본 논문에서는 제안된 방법을 검증하기 위해 데이터 시각화 시스템의 프로토타입을 개발하였다. 129 개의 웨이퍼를 에칭 공정을 할 수 있는 반도체 장비인 LAM 9600 Metal Etcher로부터 대용량 반도체 설비 데이터 세트[14]를 사용하였다. 대용량 설비 데이터를 시각화하기 위해 비주얼 아트의 개념

을 손쉽게 적용하여 코딩할 수 있는 스케치북 형태인 프로세싱 언어[15]를 사용하였으며 데이터 획득 단계에서 얻어진 대용량 설비 데이터를 효율적으로 처리하기 위해 CSV(Comma Separate Value) 파일 형식을 사용하였다.

반도체 설비로부터 연속된 순서로 획득되는 대용량 설비 데이터를 시각화하기 위해 주요 인자들의 합, 평균 및 편차와 같은 통계 분석을 사용하여 데이터를 관찰할 수 있게 하였고 스펙트럼 기반 색상을 표현할 수 있는 HSV 색상 모델을 사용하였다.

본 논문에서는 데이터의 시각화를 테스트하기 위해 <그림 7>과 같이 히트맵(Heatmap) 기반 데이터 시각화 방법을 통해 연속적인 시간에 따라 변화하는 주요 인자의 챔버별 상황을 시각화하였다. 시간에 따른 주요 인자의 이상 여부를 확인하기 위해 칼라 스케일을 반영하여 칼라 매핑을 하였다. 이상 여부를 표현하는 칼라 매핑을 위해 선택된 주요 인자에서 챔버들의 상황을 현재 인자값으로 표현하는 방법과 챔버들에 설정된 기준값과 현재 인자값과의 차이로 표현하는 방법을 사용하였다.

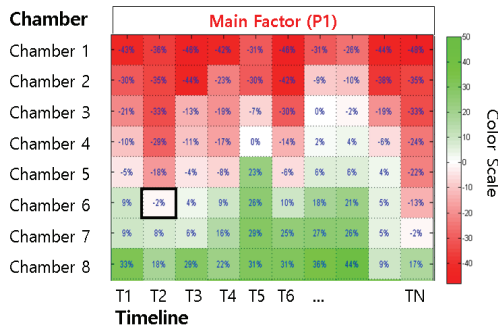


그림 7. 히트맵 기반 데이터 시각화
Figure 7. Heatmap based Data Visualization

히트맵 기반 데이터 시각화 방법과 함께 <그림 8>과 같이 데이터 시각화 매핑 및 렌더링을 위해 박스 형태(가로 축: 시간, 세로 축: 주요 인자값) 표

현, 원 형태(각도: 주요인자 값, 반지름: 시간) 표현, 데이터의 흐름을 표시할 수 있는 스트림 형태(스트림: 시간에 따라 변하는 주요인자, 스트림의 너비: 해당 시간 단계의 주요인자 값, 전체 스트림의 너비: 개별 시간 단계에서의 주요인자 값들의 총합) 표현을 사용하였다.

본 연구에서는 실시간 관찰 및 오류 감지를 위해 색상모델 HSV를 이용하여 색상(Hue)은 주요인자, 채도(Saturation)는 현재까지의 해당 인자의 평균, 밝기(Value)는 최대 밝기 값에서 해당 인자의 평균과 현재 값의 편차를 뺀 값으로 표현하였다. 이러한 데이터 시각화를 통해 오류가 발생할 가능성이 높은 구간을 실시간으로 관찰할 수 있다. 오류가 발생할 가능성이 높은 구간은 박스 형태와 원 형태 표현인 경우에는 밝기 값이 현저히 차이가 나는 구간을, 스트림 형태의 표현인 경우에는 전체 스트림의 폭이 넓은 구간을 의미한다.

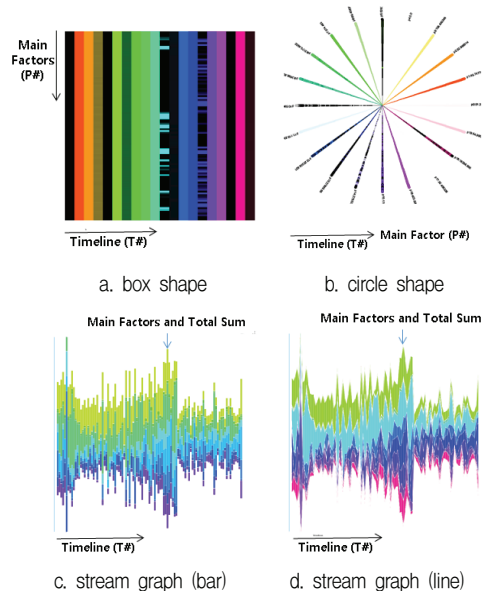


그림 8. 제안된 대화형 데이터 시각화의 결과 영상
Figure 8. The results of suggested interactive data visualization

5. 결론 및 향후 연구

본 논문에서는 설비데이터로부터 연속적으로 생성된 원시 데이터를 사용하여 데이터 시각화 기반 실시간 모니터링 및 오류 감지 시스템을 제안하였다. 데이터 시각화 기반 오류 감지 시스템은 데이터 수집, 데이터 시각화, 오류 감지 및 알람 단계로 구성하였다. 제안된 방법을 확인하기 위해 데이터 시각화 시스템의 프로토타입을 개발하였다. 설비로부터 연속적으로 생성되는 대용량 데이터를 모니터링 하기 위해 박스 형태, 원 형태, 스트림 형태 기반 데이터 시각화 매핑 시스템을 구현하였다. 이를 통해 다양한 항목의 데이터 세트를 효과적으로 한 눈에 관측할 수 있게 되었다.

제안된 방법은 이미지 생성, 데이터 융합, 오류 감지 및 기계 학습과 같은 다양한 분야에서 사용될 수 있다. 향후 연구로는 데이터 시각화한 개별 결과 이미지를 시간 순으로 적층하여 임의의 단면도 기반 데이터의 오류를 관찰 할 수 있는 블룸렌더링 기술의 활용이 필요하다. 또한, 대용량 데이터를 직접 사용하여 기계학습을 적용하는 것이 아니라 데이터 시각화 기반으로 생성된 개별 결과 영상을 사용하여 기계학습에 적용하도록 하는 연구가 필요하다.

References

- [1] *Data, data everywhere*, The Economist, Feb. 2010.
- [2] V. Dhar, *Data science and prediction*, Communications of the ACM, Vol. 56, No. 12, pp. 64, 2013.
- [3] M. Aparicio, and C. J. Costa, *Data visualization*, Communication Design Quarterly Review, Vol. 3, No. 1, pp. 7-11, Nov. 2014.
- [4] M. Friendly, *A brief history of data visualization*, Springer-Verlag, Nov. 2017.
- [5] L. J. Rosenblum, *Scientific visualization: Advances and challenges*, Academic Press, 1994.
- [6] B. B. Bederson, and B. Shneiderman, *The craft of information visualization: Readings and reflections*, Morgan Kaufmann, 2003.
- [7] P. C. Wong, and J. Thomas, *Visual analytics*, IEEE Computer Graphics and Applications, Vol. 24, No. 5, pp. 20-21, 2004.
- [8] D. A. Freedman, *Statistical models: Theory and practice*, Cambridge University Press, 2005.
- [9] H. Abdi and L. J. Williams, *Principal component analysis*, Wiley Interdisciplinary Reviews: Computational Statistics, Vol. 2, No. 4, pp. 433-459, 2010.
- [10] L. Wilkinson, and M. Friendly, *The history of the cluster heat map*, The American Statistician, Vol. 63, No. 2, pp. 179-184, May 2009.
- [11] *Find content gaps using radar charts*, Content Strategy Workshops, Mar. 2015.
- [12] L. Byron, and M. Wattenberg, *Stacked graphs-geometry & aesthetics*, IEEE Transactions on Visualization and Computer Graphics, Vol. 14, No. 6, pp. 1245-1252, 2008.
- [13] D. J. Wheeler, and D. S. Chambers, *Understanding statistical process control*, SPC Press, 1992.
- [14] B. M. Wise, N. B. Gallagher, S. W. Butler, D. D. White Jr, and G. G. Barna, *A comparison of principal components analysis, multi-way principal components analysis, tri-linear decomposition and parallel factor*

analysis for fault detection in a semiconductor etch process, Journal of Chemometrics, No. 13, pp. 379-396, 1999.

[15] D. Shiffman, *Learning processing, second edition: A beginner's guide to programming images, animation, and interaction*, Morgan Kaufmann, Aug. 2015.

대화형 데이터 시각화 기반 실시간 관측 및 오류 감지 시스템

류승택

한신대학교 컴퓨터공학부 교수

요 약

다양한 형태의 설비로부터 발생하는 여러 종류의 데이터들을 효과적으로 표현하는 방법은 쉬운 작업이 아니다. 정교한 데이터 분석 등을 위해 수집하는 데이터 항목이 많아질수록 시스템 운영자가 중요한 항목을 식별하는데 장애요소로 작용한다. 이러한 문제점을 개선하기 위해서는 대용량 데이터를 효과적으로 한 눈에 관측하고 미리 오류 감지를 확인할 수 있는 실시간 대화형 데이터 시각화 기술 개발이 필요하다. 본 논문에서는 연속적으로 발생하는 원시 데이터를 효율적으로 관리하기 위해 인터랙티브 데이터 시각화 방법을 이용한 실시간 관측 및 오류 감지 시스템을 제안한다. 제안된 시스템은 데이터 수집, 데이터 가시화, 데이터 오류 감지 및 알람 단계로 구성된다. 데이터 수집 단계에서는 데이터 수집 장치를 이용하여 발생된 실시간 설비 데이터를 저장하고 데이터 전처리과정을 통해 데이터를 마사징하고 일반화한다. 이렇게 정규화된 데이터들을 데이터 시각화 단계에서 데이터 분석, 시각화 매핑 및 렌더링 과정을 거친다. 본 논문에서는 정의된 데이터 프리미티브를 시각화 렌더링하기 위한 방법으로 히트맵(Heat Map), 레더 차트(Radar Chart), 스트림 그래프(Stream Graph) 방법을 사용하였다. 또한, 데이터의 오류 감지를 하기 위해 통계적 공정 관리(SPC: Statistical Process Control) 규칙 기반 오류 감지 및 알람 방법을 사용하였다. 제안된 방법은 이미지 생성, 데이터 융합, 오류 감지 및 기계 학습과 같은 다양한 분야에서 널리 활용될 수 있다.

감사의 글

이 논문은 한신대학교 학술연구비 지원에 의하여 연구되었음



Seung Taek Ryou received the B.S. and M.S. degrees in Computer Science Engineering from ChungAng University in 1996 and 1998, respectively. He had received his Ph.D.

in Department of Image Engineering, Graduate School of Advanced Imaging Science, Multimedia and Film from ChungAng University in 2002. He is currently an associate professor in School of Computer Engineering, Hanshin University, Korea. His interests are computer graphics, data visualization, realtime rendering, non-photorealistic rendering.

E-mail address: stryoo@hs.ac.kr