

# Visualization Tool of Distortion-Free Time-Series Matching

Seongwoo Moon<sup>†</sup> · Sanghun Lee<sup>\*\*</sup> · Bum-Soo Kim<sup>\*\*\*</sup> · Yang-Sae Moon<sup>\*\*\*\*</sup>

## ABSTRACT

In this paper we propose a visualization tool for distortion-free time-series matching. Supporting distortion-free is a very important factor in time-series matching to get more accurate matching results. In this paper, we visualize the result of time-series matching, which removes various time-series distortions such as noise, offset translation, amplitude scaling, and linear trend by using moving average, normalization, linear detrending transformations, respectively. The proposed visualization tool works as a client-server model. The client sends a user-selected time-series, of which distortions are removed, to the server and visualizes the matching results. The server efficiently performs the distortion-free time-series matching on the multi-dimensional R\*-tree index. By visualizing the matching result as five different charts, we can more easily and more intuitively understand the matching result.

**Keywords :** Visualization, Time-Series, Distortion-Free, Time-Series Matching, Index Search

## 왜곡 제거 시계열 매칭의 시각화 도구

문성우<sup>†</sup> · 이상훈<sup>\*\*</sup> · 김범수<sup>\*\*\*</sup> · 문양세<sup>\*\*\*\*</sup>

### 요약

본 논문에서는 왜곡 제거 시계열 매칭의 시각화 도구를 제안한다. 시계열 매칭에서 왜곡 제거 지원은 정확한 유사 시계열 검색을 위한 매우 중요한 요소이다. 본 논문에서는 시계열에서 발생하는 잡음, 위치 이동, 진폭 조정, 선형 추세의 왜곡들을 이동평균 변환, 정규화 변환, 선형 추세 제거 기법을 사용하여 제거한 후, 시계열 매칭을 수행하고 그 결과를 시각화한다. 본 논문에서 제안하는 시각화 도구는 클라이언트-서버 모델을 기반으로 구현된다. 클라이언트는 사용자가 선택한 시계열의 왜곡을 제거하여 서버에 전달하고, 매칭 결과의 시각화를 수행한다. 서버는 R\*-트리 기반의 다차원 인덱스로 왜곡 제거 시계열 매칭을 효율적으로 수행한다. 왜곡 제거 시계열 매칭의 결과는 다섯 가지 차트로 달리 표현하여 사용자가 직관적이고 쉽게 이해할 수 있도록 구현한다.

**키워드 :** 시각화, 시계열, 왜곡 제거, 시계열 매칭, 인덱스 검색

### 1. 서론

시각화는 데이터 혹은 질의 결과를 이미지, 차트, 다이어그램 등의 시각적 도구로 표현하는 방법이다. 이 같은 시각화는 의미 있는 정보를 시각적으로 전달할 수 있어 과학, 교육, 정보 등 여러 분야에서 사용되고 있다[1]. 특히, 데이터마이닝 분야에서는 SNA(Social Network Analysis), 시계열, 이미지 매칭 등을 효과적으로 표현하여 사용자의 데이

터 인식과 이해, 분석 및 처리의 효율성을 향상시키는 데 널리 사용되고 있다[1].

사물인터넷(IoT) 환경에서는 시계열 데이터가 많이 발생하고 동시에 왜곡 또한 발생한다. 특히, 심전도와 뇌파 같은 의료 관련 시계열 데이터는 정확한 데이터 측정을 위해 왜곡 제거 지원이 반드시 필요하다.

본 논문에서는 왜곡 제거를 지원하는 시계열 매칭의 시각화 도구를 설계하고 구현한다. 왜곡 제거 시계열 매칭은 시계열에 나타나는 왜곡을 제거하여 데이터베이스의 유사 시계열을 검색하는 방법으로, 보다 정확하고 직관적인 유사 시계열 매칭이 가능하다. 시계열에 발생하는 왜곡은 크게 잡음(noise), 위치 이동(offset translation), 진폭 조정(amplitude scaling), 선형 추세(linear trend)가 있다[2-3]. 본 논문에서는 클라이언트-서버 모델에서 이들 왜곡을 제거하는 효율적인 시계열 매칭을 설계한다. 구체적으로는 클라이언트와 서버의 시스템 구조를

※ 이 논문은 2014년도 정부(미래창조과학부)의 지원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2014R1A2A2A01002548).

※ 2014년도 강원대학교 학술연구조성비로 연구하였음(과제번호-C1010859-01-01).

† 준회원: 강원대학교 컴퓨터과학과 석사과정

\*\* 준회원: 강원대학교 컴퓨터과학과 박사과정

\*\*\* 정회원: 한국과학기술원 산업경영연구소 연수연구원

\*\*\*\* 종신회원: 강원대학교 컴퓨터과학과 교수

Manuscript Received: July 29, 2015

First Revision: September 1, 2015

Accepted: September 3, 2015

\*Corresponding Author: Yang-Sae Moon(ysmoon@kangwon.ac.kr)

제시하고, 두 시스템의 동작 내용을 알고리즘으로 제시한다. 또한, 구현 결과의 시각화 내용을 사례를 들어 설명한다.

왜곡 제거를 위해 기존의 다양한 왜곡 제거 기법들을 활용하였다. 먼저, 잡음은 Moon 등[4]과 Loh 등[5]의 이동평균 변환을 이용하여 해결하였고, 위치 이동과 진폭 조정은 Moon 등[6]과 Loh 등[7]의 정규화 변환을 이용하여 해결하였다. 또한 선형 추세는 Gil 등[8]의 최소 제곱법을 사용하여 해결하였다. 구현 결과, 제안한 왜곡 제거 매칭 시각화 도구는 왜곡 제거 기법을 적용한 후 질의 시계열과 유사한 데이터 시계열을 검색하여 시각적으로 사용자에게 제공하였다. 특히, 검색 결과를 다섯 종류의 차트(Line, Range, k-NN, Point, Polar)를 통해 시각화하여 보다 쉽고 직관적으로 시계열 매칭 결과를 비교할 수 있다.

본 논문의 구성은 다음과 같다. 2절은 관련 연구로 시계열 왜곡의 네 가지 종류를 설명하고 기존의 시계열 시각화 연구를 소개한다. 3절은 본 논문에서 사용하는 세 가지 시계열 왜곡 제거 기법을 구체적으로 설명한다. 4절에서는 클라이언트-서버 모델로 구성된 시각화 프레임워크를 각 시스템 구조와 알고리즘을 통해 설명하고, 왜곡 제거 시계열 매칭 수행 결과를 다섯 가지 차트로 소개한다. 마지막으로, 5절에서 결론을 맺는다.

2. 관련 연구

시계열은 연속적인 데이터를 일정 간격으로 측정한 실수 값의 시퀀스이다[9-12]. 시계열 데이터의 예로 주식 데이터, 환율 데이터, 의료 데이터 등이 있다. 시계열 매칭은 사용자에 의해 주어진 질의 시퀀스(질의 시계열)와 데이터베이스에 저장된 데이터 시퀀스(데이터 시계열)를 비교하여 유사한 데이터 시퀀스를 찾는 작업이다[9, 13]. 본 논문에서는 시계열 매칭 결과, 질의 시계열과 가장 유사한 k개의 데이터 시계열을 찾는 k-NN 질의[14-15]를 사용하고, 유사 거리 함수로는 유클리디안(Euclidean) 거리 모델[3, 4, 16]을 사용한다. 길이 n인 두 시계열  $X=(\{X[1], X[2], \dots, X[n]\})$ 와  $Y=(\{Y[1], Y[2], \dots, Y[n]\})$ 의 유클리디안 거리  $D(X,Y)$ 는 Equation (1)과 같이 정의된다.

$$D(X, Y) = \sqrt{\sum_{i=1}^n |X[i] - Y[i]|^2} \quad (1)$$

시계열에서 발생하는 왜곡의 종류는 Fig. 1과 같이 잡음, 위치 이동, 진폭 조정, 선형 추세의 네 가지로 나눌 수 있다[3]. Fig. 1(A)에서 보듯이, 잡음은 시계열에 백색 잡음(white noise)이 존재하는 경우를 뜻하며, 이러한 잡음으로 인한 시계열 왜곡은 이동평균 변환으로 제거할 수 있다. Fig. 1(B)의 위치 이동은 모양이 유사한 두 시계열이 y축을 기준으로 다른 위치에 존재하는 경우를 의미하고, Fig. 1(C)의 진폭 조정은 두 시계열의 진폭이 서로 다른 경우를 뜻한다. 위치

이동과 진폭 조정은 정규화 변환을 이용하여 왜곡을 해결할 수 있다. 마지막으로, 선형 추세는 Fig. 1(D)와 같이 유사한 두 시계열이 서로 다른 직선에 비례하여 증가/감소하는 추세를 갖는 경우를 의미하며, 선형 추세 제거법으로 해결이 가능하다[3, 8].

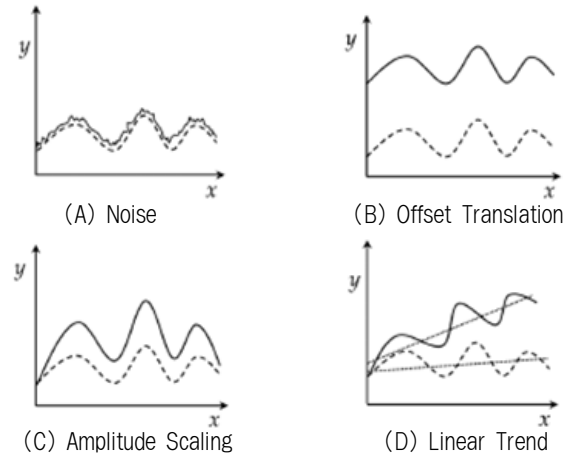
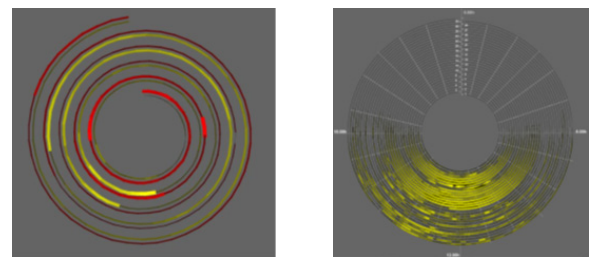


Fig. 1. Type of Distortion which Can be Displayed in Time-Series

최근 시계열 매칭의 시각화 연구에서 Lee 등[17]은 시계열 매칭에 대한 사전지식이 없는 사용자를 위해 시계열 매칭 과정과 결과를 시각적으로 제공하였다. 구현된 시각화 도구는 데이터 시계열과 질의 시계열의 비교 및 DTW (Dynamic Time Warping)[18] 거리 함수의 계산 행렬을 시각적으로 표현한다. 다른 시계열 시각화 방법으로는 Fig. 2에서 보듯이 Weber 등[19]의 나선형 기반의 시계열 시각화 방법이 있다. 이 방법은 시계열 데이터의 주기를 나선형으로 표시하고, 시계열 값의 크기를 나선상 색의 농도나 선의 굵기 등으로 표현하여 데이터를 쉽게 알 수 있게 하였다. 그러나 앞서 언급한 두 방법은 모두 순수 시계열 데이터 자체의 시각화에 초점을 맞추어 구현되었으며, 왜곡된 시계열의 유사 매칭 결과를 비교 및 분석하는 기능을 지원하지 않는다. 반면에, 본 논문에서 제안하는 시각화 도구는 왜곡 제거를 지원하고, 다섯 가지 차트로 시계열 매칭 결과를 시각화하여 다양한 비교 및 분석이 가능하다.



(A) Stock Prices of Microsoft (Yellow) and Sun Microsystems (Red) in Five Years (B) Sunshine Intensity (Yellow) and Sun Microsystems Using About the Same Screen Real Estate

Fig. 2. Spiral Time-series Visualization of Weber[19]

### 3. 왜곡 제거 종류와 매칭

본 절에서는 시각화 도구의 구현에 사용되는 기본(naive) 및 왜곡 제거 시계열 매칭 기법들을 소개한다. 먼저, 기본 매칭은 시계열 매칭의 가장 기본적인 매칭 방법으로서, 왜곡되지 않은 순수(raw) 시계열 데이터를 사용하여 시계열 매칭을 수행한다. 기본 매칭은 구현과 이해가 쉬워 가장 널리 사용되는 매칭 방법이지만, 시계열에 대한 적절한 왜곡 제거 방법을 지원하지 않기 때문에 앞서 Fig. 1에서 살펴본 왜곡에 취약하다는 단점이 있다.

이동평균 변환 매칭은 시계열에 포함된 잡음을 제거하여 정확한 시계열 매칭의 수행을 지원하는 기법이다. 이동평균 변환[20]은 주어진 시계열에서 연속된  $k$ 개 엔트리의 평균값을 새로운 엔트리로 하는 시계열을 구성하는 변환이다. Equation (2)는 길이  $n$ 인 시계열  $S=(S[1], \dots, S[n])$ 로부터  $k$ -이동평균 변환한 길이  $n-k+1$ 인 시계열  $S^{(k)}=(S^{(k)}[1], \dots, S^{(k)}[n-k+1])$ 를 구하는 공식이다.

$$S^{(k)}[i] = \frac{1}{k}(S[i] + S[i+1] + \dots + S[i+k-1])$$

$$= \frac{1}{k} \sum_{j=i}^{i+k-1} S[j], \quad i = 1, \dots, n-k+1. \quad (2)$$

이동평균 변환에서  $k$ 값을 이동평균 계수라 하는데, 조건에 따라 잡음의 제거 정도를 달리하여 사용될 수 있다. 본 논문에서는 Fig. 1(A)의 잡음 제거를 위해 Equation (2)의 이동평균 변환을 사용한다.

정규화 변환 매칭은 위치 이동(Fig. 1(B))과 진폭 조정(Fig. 1(C))된 시계열의 왜곡을 해결하여 시계열 매칭을 수행하는 기법이다. 정규화 변환[7]은 주어진 시계열의 평균과 표준편차를 이용하여 계산된 엔트리로 새로운 시계열을 구성하는 변환이다. Equation (3)은 시계열  $S$ 로부터 정규화된 시계열  $\bar{S}$ 를 구하는 정규화 변환을 수행하는 공식이다[21].

$$\bar{S}[i] = \frac{S[i] - \mu(S)}{\sigma(S)} \quad (3)$$

Equation (3)에서,  $\mu(S)$ 와  $\sigma(S)$ 는 각각 시계열  $S$ 의 평균과 표준편차를 나타낸다. 본 논문에서는 Fig. 1(B)와 1(C)의 위치 이동과 진폭 조정의 해결을 위해 Equation (3)의 정규화 변환을 사용한다.

마지막으로, 선형 추세는 시계열에 나타나는 왜곡 중 하나로 선형 추세를 나타내는 직선을 구하는 방법은 여러 가지가 있다. 본 논문에서는 추세를 구하기 위해 가장 대표적인 방법인 최소 제곱법을 사용한다. 시계열  $S=(S[1], S[2], \dots, S[n])$ 에 대한 직선 방정식의 기울기  $\alpha$ 와 절편  $\beta$ 를 구하는 최소 제곱법은 Equation (4)와 같다[3].

$$\alpha = \frac{n \sum_{i=1}^n i S[i] - \sum_{i=1}^n i \sum_{i=1}^n S[i]}{n \sum_{i=1}^n i^2 - (\sum_{i=1}^n i)^2},$$

$$\beta = \frac{\sum_{i=1}^n S[i]}{n} - \alpha \frac{\sum_{i=1}^n i}{n}, \quad i = 1, \dots, n. \quad (4)$$

따라서, 선형 추세에 대한 직선의 방정식  $g(i) = \alpha i + \beta$ 는 Equation (4)에서 구한 기울기  $\alpha$ 와 절편  $\beta$ 를 각각 대입하여 구할 수 있다. 선형 추세 제거[3]는 원본 시계열  $S$ 에서 해당 선형 추세  $g(i) = \alpha i + \beta$ 를 제거하여 새로운 시계열  $\hat{S}$ 를 생성하는 방법을 의미하며, Equation (5)로 표현할 수 있다.

$$\hat{S}[i] = S[i] - g[i], \quad i = 1, \dots, n. \quad (5)$$

본 논문에서는 선형 추세 제거를 위해 Equation (4)와 (5)의 선형 추세 제거 시계열을 구한 후 시계열 매칭에 활용한다.

본 논문에서 제안한 시각화 도구는 앞서 설명한 왜곡 제거 기법을 각각 적용하여 사용자에게 각 왜곡 제거 시계열 매칭 결과의 차이를 시각적으로 확인할 수 있게 제공한다. 따라서, 각 왜곡 제거 기법을 하이브리드 하여 시계열 매칭을 수행하는 방법은 향후 연구로 다룰 예정이다.

### 4. 왜곡 제거 시계열 매칭 시스템의 시각화 설계 및 구현

#### 4.1 시스템 동작 프레임워크

본 논문에서 제안하는 시계열 매칭 시각화 시스템은 Fig. 3과 같이 클라이언트-서버 모델로 구성된다. 먼저, 클라이언트는 사용자로부터 질의 시계열을 입력받아 사용자가 선택한 왜곡 제거 방법을 통해 왜곡 제거 시계열을 생성하며, 질의 시계열과 왜곡 제거 시계열, 왜곡 제거 방법, 저장된 변환 방법,  $k$ -NN 질의 계수  $k$ 를 서버로 전송한다. 서버는 전달받은 데이터를 사용하여 왜곡 제거 시계열 매칭을 수행하고, 그 결과로 질의 시계열과 유사한  $k$ 개의 데이터 시계열을 클라이언트에게 반환한다. 그리고 클라이언트는 서버에서 전달받은 질의 처리 결과를 시각화한다.

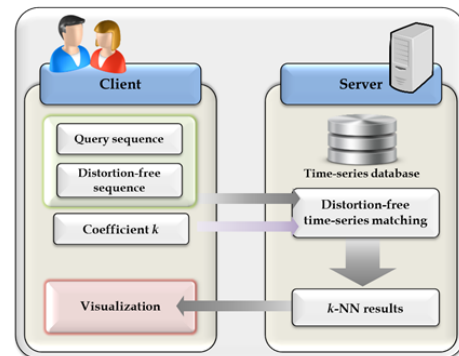


Fig. 3. Overall Framework of the Proposed Visualization System

### 4.2 클라이언트 시스템 설계

제안하는 시계열 매칭 시각화 시스템의 클라이언트는 왜곡 제거, 저차원 변환, 데이터 구조 변환, 통신, 시각화의 다섯 가지 모듈로 구성된다. 이들 다섯 모듈을 차례로 설명하면 다음과 같다. 첫째, 왜곡 제거 모듈은 사용자가 선택한 시계열을 3절에서 소개한 방법으로 왜곡 제거하는 모듈로서, 이동평균 변환, 정규화 변환, 선형 추세 제거 등을 수행하는 모듈이다. 둘째, 저차원 변환 모듈은 왜곡을 제거한 시계열을 인덱스에 사용하기 위해 저차원 변환을 수행하는 모듈이다. 저차원 변환 함수로는 DWT(Discrete Wavelet transform)[22], DCT(Discrete Cosine Transform)[23], SVD(Singular Value Decomposition)[24], DFT(Discrete Fourier Transform)[9], PAA(Piecewise Aggregate Approximation)[12] 등이 있다. 이때, 저차원 변환을 수행한 시계열과 왜곡 제거 시계열은 구조체로 구성된다. 셋째, 데이터 구조 변환 모듈은 클라이언트-서버 간 원활한 데이터 전송을 위해 앞서 구성한 구조체를 바이트 스트림으로 변환하고, 시계열 매칭 수행 후 반환된 결과를 다시 바이트 스트림에서 구조체로 변환하는 모듈이다. 넷째, 통신 모듈은 서버에 질의 구조체를 전송하고 그 결과를 받기 위한 모듈로서 클라이언트-서버 구조에서 널리 사용되는 TCP/IP 소켓 통신을 사용한다. 다섯째, 클라이언트의 주요 기능인 시각화 모듈은 서버에서 전달받은 결과를 차트로 시각화하는 모듈로서 MSChart 라이브러리[25]를 사용하여 구현되며, 자세한 구현사항은 4.4절에서 설명한다.

Fig. 4와 5는 앞서 설명한 클라이언트의 모듈을 표현한 시스템 구조도와 알고리즘을 각각 나타낸다. 알고리즘을 보면, 먼저 사용자는  $k$ -NN 질의 계수  $k$ , 왜곡 제거 방법  $d$ , 저차원 변환 방법  $l$ 을 선택한다. 다음으로, 사용자가 선택한 질의 시계열을 앞서 선택한 왜곡 제거 기법  $d$ 와 저차원 변환 기법  $l$ 을 사용하여 왜곡 제거 시계열  $q'$ 과 저차원 변환 시계열  $q''$ 를 생성한다. Fig. 5의 라인 (1)과 (2)가 이와 같이 시계열 생성을 수행하는 과정이다. 이때, 사용자 선택 값  $k$ ,  $d$ ,  $l$ 과 시계열  $q'$ ,  $q''$ 를 구조체로 구성하고, 구조체 변환 모듈에서 질의 구조체를 바이트 스트림으로 변환 후 통신 모듈을 이용하여 질의 값을 서버에 전달한다. Fig. 5의 라인 (3)과 (4)가 이 같은 질의 변환과 전송 작업에 해당한다. 마지막으로, 서버에서 반환된 왜곡 제거 시계열 매칭 결과 값을 시각화 모듈을 통해 시각화를 수행한다. Fig. 5의 라인 (5)가 서버에서 반환받은 결과 값  $R$ 을 사용하여 시각화를 수행하는 과정이다.

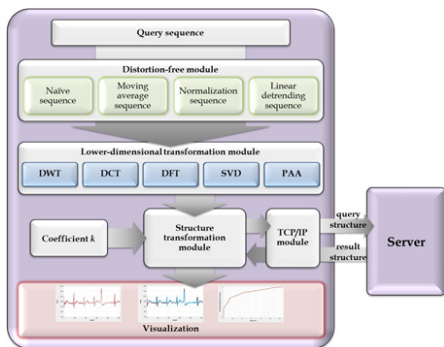


Fig. 4. Operating Structure of Client System

### Algorithm Client

```

Input:
  q: a query sequence given by a user;
  d: distortion type;
  l: lower-dimensional transformation type;
  k: the number of resulting data sequences;
Output:
  Visualize k similar data sequences;
begin
  (1) q' := Remove_Distortion(q, d);
  (2) q'' := Perform_Dimensionality_Reduction(q', l);
  (3) S := Construct_Byte_Stream(q', q'', d, l, k);
  (4) R := Server(S); //Send S to Server and get the result R
  (5) Visualize(R); //Visualize k sequences of R
end
    
```

Fig 5. Client Algorithm

### 4.3 서버 시스템 설계

본 시각화 시스템의 서버는 시계열 매칭, 데이터 입출력, 인덱스의 세 가지 모듈로 구성되어 있다. 첫째, 인덱스 모듈은 다차원 인덱스를 구성하는 모듈로서, 다차원 인덱스 구성 과정은 전처리로 수행되며 데이터베이스의 시계열을 3절의 세 가지 왜곡 제거 모듈을 통해 왜곡을 제거하고, 각각 다섯 가지의 저차원 변환을 수행하여 총 15가지의 R\*-트리[11]를 미리 생성한다. 둘째, 시계열 매칭 모듈은 클라이언트에서 전달받은 시계열들을 사용하여  $k$ -NN 질의를 수행하는 모듈이다. 셋째, 데이터 입출력 모듈은 클라이언트의 데이터 구조 변환 모듈과 같이 통신에 사용하는 바이트 스트림과 서버 구조체 사이의 변환을 수행하는 모듈이다.

Fig. 6과 7은 앞서 설명한 서버의 모듈들을 표현한 시스템 구조도와 알고리즘이다. 먼저, 서버는 전처리 모듈을 사용하여 다차원 인덱스를 구성한다(전처리 참조). 다음으로, 클라이언트에서 전달받은 질의 시계열을 왜곡 제거 시계열 매칭 모듈에서 다차원 인덱스 검색을 수행한다(라인 (1)). 마지막으로, 검색된 결과의 유효성 판단을 위한 후처리로 시계열 데이터베이스 검색을 수행하고(라인 (2)), 유효한 결과 데이터를 클라이언트에 반환한다(라인 (3)).

### 4.4 왜곡 제거 시계열 매칭의 시각화 구현

제안한 시각화 도구는 클라이언트-서버 모델을 사용하였으며, 구현 및 실험 환경은 다음과 같다. 클라이언트는 Windows 7 운영체제에서 Microsoft Visual Studio 2010을 사용하여 C#과 MSChart를 기반으로 구현하였고, 서버는 Cent-OS 5.9 Linux 운영체제에서 GNU C를 사용하여 구현하였다. 실험에 사용한 데이터는 ECG 데이터[26]로, 1024 차원으로 구성된 200,000개의 데이터를 사용하였고, 이 중

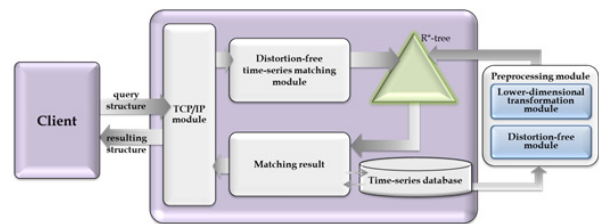


Fig. 6. Operating Structure of Server System

**Algorithm Server**

Preprocessing step: multi-dimensional indexes are constructed prior to the matching process.

Input:

**S**: the data structure received from the client;  $//(q', qf, d, l, k)$   
**k**: the number of data sequences to be returned;

Output:

**R**: a set of matching results;  $//(d_x, d(q, d_x))$

**Begin**

- (1)  $Select\_Index(d, l);$
- (2)  $R' := Search\_Index(q', qf, k);$
- (3)  $R := Post\_Processing(R');$
- (4) return **R**;

**end**

Fig. 7. Server Algorithm

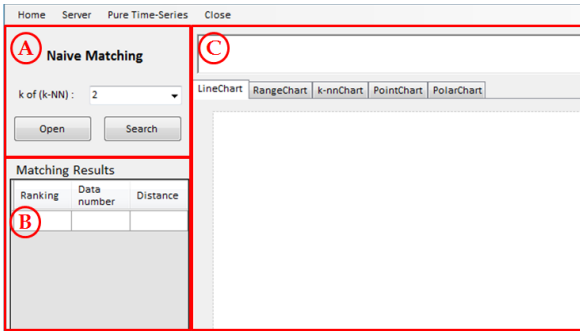


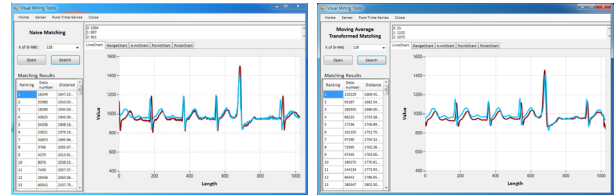
Fig. 8. Initial Screen of the Client Visualization Tool

19,500개를 질의 데이터로 사용하였다. 또한, 4.2절에서 언급한 다섯 가지의 저차원 변환 함수를 활용하여 각각 다차원 인덱스를 구성하였다. 본 논문의 시각화 실험에서는 저차원 변환 함수 가운데 PAA를 차원 수 8로 고정하여 사용하였다.

Fig. 8은 왜곡 제거 시계열 매칭 시각화 도구의 초기화면이다. 화면에서 ① 부분의 “Open” 버튼은 질의 시계열을 선택하여 왜곡 제거 시계열과 저차원 변환 시계열을 생성하고, “Search” 버튼은 왜곡 제거 시계열 매칭을 통해 질의 시계열과 유사한 시계열을 검색한다. 또한, “k of (k-NN)”은 질의 결과로 돌려받을 k개의 데이터 시계열을 선택하는 부분이다. 다음으로, ② 부분은 리스트를 이용하여 k개의 질의 결과를 사용자에게 보여주는 부분으로, 리스트의 각 열은 질의 시계열과 유사한 데이터 시계열의 순위, 시계열 번호, 그리고 유사거리 값을 각각 나타낸다. 마지막으로, ③ 부분은 본 도구의 왜곡 제거 시계열 매칭 결과를 시각화하여 사용자에게 보여주는 부분이다.

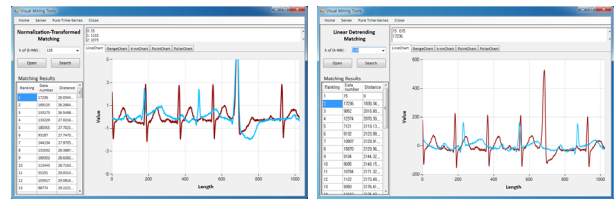
Fig. 9는 Fig. 8의 실행 결과로서 기본(naive) 및 왜곡 제거 매칭 결과를 Line 차트로 나타낸 화면이다. 먼저, Fig. 9(A)는 기본 매칭 결과를 시각화한 화면으로 왜곡이 제거되지 않은 일반적인 시계열 매칭 결과를 나타낸다. 다음으로, Fig. 9(B)는 이동평균 변환 매칭 결과를 시각화한 화면으로, Fig. 9(A)와 비교하면 시계열의 잡음이 제거된 모습을 확인할 수 있다. Fig. 9(C)는 정규화 변환 매칭 결과를 시각화한 화면으로 차트의 세로축이 정규화되어 나타난다. 마지막으로, Fig. 9(D)는 선형 추세 제거 매칭 결과를 시각화한 화면으로, 질의 및 데이터 시계열의 추세가 제거된 상태로 나타

난다. 이 실험에서는 Fig. 9(A)와 9(B)의 결과가 동일하게 나타났으나, Fig. 9(C)와 9(D)는 다른 결과가 도출되었다. 이는 왜곡 제거에 의해 기본 매칭과는 다른 시계열이 검색될 수 있음을 의미한다.



(A) Naive Matching

(B) Moving Average Transformation Matching



(C) Normalization

(D) Linear Detrending Matching

Transformation Matching

Fig. 9. Screenshots of Distortion-Free Time-Series Matching Result

Fig. 10은 시각화 도구의 차트 종류로서 각각 Range, k-NN, Point, Polarpoint 차트로 시각화한 화면이다. 먼저, Range 차트(Fig. 10(A))는 질의 시계열과 데이터 시계열 사이의 차이를 표시한 차트로, 두 시계열 사이의 거리를 차트의 면적을 통해 직관적으로 확인할 수 있다. 다음으로, k-NN 차트(Fig. 10(B))는 k-NN 질의 결과 검색된 데이터 시계열의 순위별 유사거리를 시각화한 화면이다. 그림과 같이, k-NN 차트의 기울기를 통해 데이터 시계열 사이의 상대적 유사 정도를 직관적으로 확인할 수 있고, 기울기의 크기를 통해 결과 데이터 중 몇 개를 선택할지 판단할 수 있다. 마지막으로, Fig. 10(C)와 10(D)는 데이터 시계열을 점으로 표현한 차트이다. Fig. 10(C)는 k-NN 질의 결과로 검색된 k개의 데이터 시계열들을 Point 차트에 표현한 것이다. 차트의 x축은 시계열의 ID를 나타내고 x축의 범위는 데이터베이스에 저장된 전체 시계열의 개수이며, y축은 질의 시계열과의 거리를 나타낸다. 이 차트를 통해 검색된 유사 시계열들이 전체 데이터베이스 상에 밀집된 위치와 밀집 정도를 직관적으로 확인할 수 있다. Fig. 10(D)는 Fig. 10(C)에서 구현한 차트를 Polar 차트로 재구성한 것이다. Polar 차트의 원둘레는 Fig. 10(C)의 x축과 동일하고, 반지름은 Fig. 10(C)의 y축과 동일하다. 차트가 원형으로 그려지기 때문에 데이터 시계열들의 밀집 정도를 Point 차트와는 다른 시각으로 확인할 수 있다.

Fig. 11은 Fig. 10의 정규화 변환 매칭의 결과를 네 가지 차트로 시각화한 화면이다. 각 그림에서 보듯이, 동일한 질의 시계열을 선택하더라도 왜곡 제거 매칭 방법에 따라 그 결과

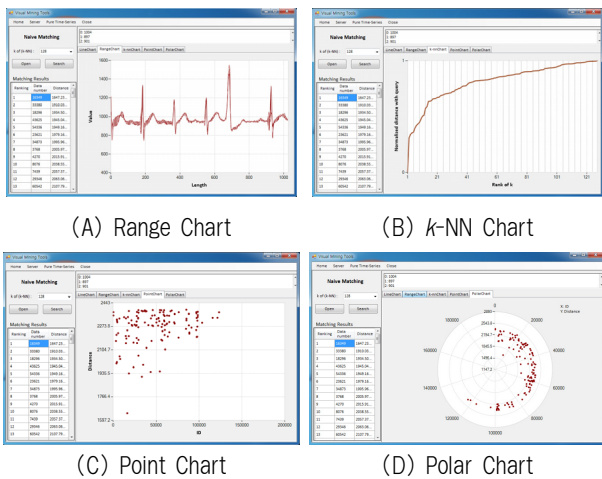


Fig. 10. Chart Type of Naive Matching

가 다를 수 있다. 정규화 변환 이외에 이동 평균, 선형 추세 제거에 대해서도 네 가지 종류의 차트로 시각화할 수 있다.

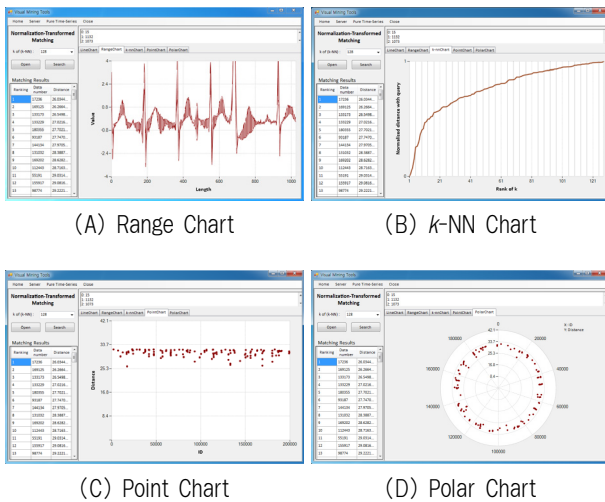


Fig. 11. Chart Type of Normalization Transformation Matching

### 5. 결 론

본 논문은 왜곡 제거 시계열 매칭 결과를 여러 종류의 차트를 사용하여 시각화하였다. 설계 및 구현한 시스템은 클라이언트-서버 구조를 가지며, 본 논문에서는 클라이언트와 서버의 각 시스템 구조와 알고리즘을 제시하였다. 제한한 시스템은 시계열에 나타날 수 있는 네 종류의 왜곡들을 이동평균 변환, 정규화 변환, 선형 추세 제거 기법을 통해 제거하였으며, 시계열 매칭 결과를 다섯 종류의 차트를 사용하여 시각화하였다. 첫째, Line 차트를 활용하여 기본 매칭과 왜곡 제거 매칭 간의 차이를 시각적으로 쉽게 확인하였으며 매칭 결과를 직관적으로 비교하였다. 둘째, Range 차

트를 통해 두 시계열 간의 거리를 면적으로 표현하여 시계열 간의 유사도를 직관적으로 비교하였다. 셋째, k-NN 차트의 기울기를 통해 결과 데이터 사이의 상대적 유사 정도를 확인하고, 유효한 데이터 선택의 판단을 가능케 하였다. 넷째, Point 차트와 Polar 차트를 사용하여 결과 시계열이 전체 데이터베이스에서 밀집된 위치와 정도를 직관적으로 확인하였다. 향후 연구로는 심전도, 기후, 이미지 데이터와 같은 다양한 멀티미디어 데이터의 시각화와 실시간으로 시계열 데이터를 분석할 수 있는 시각화를 다룰 예정이다.

### References

- [1] Visualization [Internet], <http://en.wikipedia.org/wiki/Visualization/>.
- [2] E. Keogh and C. A. Ratanamahatana, "Indexing and Mining Large Time Series Databases," in *Proc. of 12th Int'l Conf. on Database Systems for Advanced Applications*, Tutorial, Bangkok, Thailand, Apr., 2007.
- [3] M.-S. Gil, B.-S. Kim, M.-J. Choi, and Y.-S. Moon, "Fast Index Construction in Distortion-Free Time-Series Subsequence Matching," in *Proc. of the 2nd Int'l Conf. on Big Data and Smart Computing*, Jeju, Korea, pp.130-135, Feb., 2015.
- [4] Y.-S. Moon and J. Kim, "Efficient Moving Average Transform-Based Subsequence Matching Algorithms in Time-Series Databases," *Information Sciences*, Vol.177, No.23, pp.5415-5431, Dec., 2007.
- [5] W.-K. Loh, S.-W. Kim, and K.-Y. Whang, "Index Interpolation: A Subsequence Matching Algorithm Supporting Moving Average Transform of Arbitrary Order in Time-Series Databases," *IEICE Trans. on Information and Systems*, Vol.E84-D, No.1, pp.76-86, Jan., 2001.
- [6] Y.-S. Moon and J. Kim, "Fast Normalization-Transformed Subsequence Matching in Time-Series Databases," *IEICE Trans. on Information and Systems*, Vol.E90-D, No.12, pp. 2007-2018, Dec., 2007.
- [7] W.-K. Loh, S.-W. Kim, and K.-Y. Whang, "A Subsequence Matching Algorithm that Supports Normalization Transform in Time-Series Databases," *Data Mining and Knowledge Discovery*, Vol.9, No.1, pp.5-28, Jul., 2004.
- [8] M.-S. Gil, Y.-S. Moon, and B.-S. Kim, "Linear Detrending Subsequence Matching in Time-Series Databases," *IEICE Trans. on Information and Systems*, Vol.E94-D, No.4, pp. 917-920, Apr., 2011.
- [9] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient Similarity Search in Sequence Databases," in *Proc. of the 4th Int'l Conf. on Foundations of Data Organization and Algorithms*, Chicago, Illinois, pp.69-84, Oct., 1993.
- [10] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast Subsequence Matching in Time-Series Databases," in *Proc. of Int'l Conf. on Management of Data*, ACM SIGMOD,

Minneapolis, Minnesota, pp.419-429, May, 1994.

[11] Y.-S. Moon and B. S. Lee, "Safe MBR-Transformation in Similar Sequence Matching," *Information Sciences*, Vol.270, pp.28-40, Jun., 2014.

[12] Y.-S. Moon, B.-S. Kim, M. S. Kim, and K.-Y. Whang, "Scaling-invariant Boundary Image Matching Using Time-series Matching Techniques," *Data & Knowledge Engineering*, Vol.69, No.10, pp.1022-1042, Oct., 2010.

[13] S. Lee, B.-S. Kim, M.-J. Choi, and Y.-S. Moon, "An approximate Multi-Step k-NN Search in Time-Series Databases," *Advances in Computer Science and its Applications*, Vol.279, pp.173-178, Dec., 2013.

[14] W.-S. Han, J. Lee, Y.-S. Moon, and H. Jiang, "Ranked Subsequence Matching in Time-Series Databases," in *Proc. of the 33rd Int'l Conf. on Very Large Data Bases*, Vienna, Austria, pp.423-434, Sep., 2007.

[15] H.-S. Lim, K.-Y. Whang, and Y.-S. Moon, "Similar Sequence Matching Supporting Variable-Length and Variable-Tolerance Continuous Queries on Time-Series Data Stream," *Information Sciences*, Vol.178, No.6, pp. 1461-1478, Mar., 2008.

[16] Y.-S. Moon, K.-Y. Whang, and W.-K. Loh, "Duality-Based Subsequence Matching in Time-Series Databases," in *Proc. of 17th Int'l Conf. on Data Engineering*, IEEE, Heidelberg, Germany, pp.263-272, Apr. 2001.

[17] S.-J. Lee, J. Lee, H. Cho, and W. S. Han, "A Visualization Tool for Ranked Subsequence Matching in Time-Series Databases," *Journal of KIISE*, Vol.38, No.2, pp.92-103, Apr., 2011.

[18] T. Rakthanmanon, et al., "Searching and Mining Trillions of Time Series Subsequences under Dynamic Time Warping," in *Proc. of the 18th Int'l Conf. on Knowledge Discovery and Data Mining*, Beijing, China, pp.262-270, Aug., 2012.

[19] M. Weber, M. Alexa, and W. Muller, "Visualizing Time-Series on Spirals," in *Proc. of Int'l Conf. on IEEE Symp. on Information Visualization*, San Diego, California, pp.7-13, Oct., 2001.

[20] B.-S. Kim, Y.-S. Moon, M.-J. Choi, and J. Kim, "Interactive Noise-Controlled Boundary Image Matching Using the Time-Series Moving Average Transform," *Multimedia Tools and Applications*, Vol.72, Issue.3, pp.2543-2571, Oct., 2014.

[21] W.-K. Loh, Y.-S. Moon, and J. Srivastava, "Distortion-free Predictive Streaming Time-Series Matching," *Information Sciences*, Vol.180, No.8, pp.1458-1476, Apr., 2010.

[22] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental Comparison of Representation Methods and Distance Measures for Time-Series Data," *Data Mining and Knowledge Discovery*, Vol.26, No.2, pp.275-309, Mar., 2013.

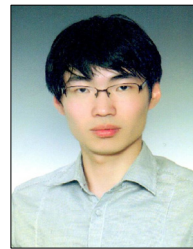
[23] S. Papadimitriou and P. S. Yu, "Optimal Multi-Scale Patterns

in Time Series Streams," in *Proc. of Int'l Conf. on Management of data*, ACM SIGMOD, Chicago, Illinois, pp.647-658, Jun., 2006.

[24] H. Chen and Y. Zhu, "A Robust Video Watermarking Algorithm based on Singular Value Decomposition and Slope-based Embedding Technique," *Multimedia Tools and Applications*, Vol.71, No.3, pp.991-1012, Aug., 2014.

[25] MSChart [Internet], <https://code.msdn.microsoft.com/mschart/>.

[26] ECG Data [Internet], <http://www.physionet.org/>.



### 문 성 우

e-mail : swmoon@kangwon.ac.kr

2014년 강원대학교 컴퓨터과학과  
(학사)

2014년~현 재 강원대학교 컴퓨터과학과  
석사과정

관심분야 : Data Mining & Knowledge  
Discovery, Data Mining  
Applications, Visualization,  
Image Processing



### 이 상 훈

e-mail : sanghun@kangwon.ac.kr

2011년 강원대학교 컴퓨터과학과(학사)

2013년 강원대학교 컴퓨터과학과(석사)

2013년~현 재 강원대학교 컴퓨터과학과  
박사과정

관심분야 : Data Mining & Knowledge  
Discovery, Data Mining  
Applications, Privacy-  
Preserving Data Mining



### 김 범 수

e-mail : bskim@m Mozart.kaist.ac.kr

2006년 한라대학교 컴퓨터공학과(학사)

2008년 강원대학교 컴퓨터과학과(석사)

2013년 강원대학교 컴퓨터과학과(박사)

2013년~2014년 한국과학기술원 첨단정보기술  
연구센터 박사후연구원

2014년~2014년 강원대학교 정보통신연구소 연수연구원

2015년~현 재 한국과학기술원 산업경영연구소 연수연구원

관심분야 : Data Mining & Knowledge Discovery,  
Image Data Mining, Data Mining Applications,  
Privacy-preserving Data Mining



## 문 양 세

e-mail : ysmoon@kangwon.ac.kr

1991년 한국과학기술원 전산학과(학사)

1993년 한국과학기술원 전산학과(석사)

2001년 한국과학기술원 전산학전공(박사)

1993년~1997년 현대전자산업(주)

주임연구원

2001년~2002년 (주)현대시스콤 선임연구원

2002년~2005년 (주)인프라밸리 기술위원(이사)

2005년~2008년 한국과학기술원 첨단정보기술연구센터 연구원

2008년~2009년 미국 퍼듀대학교 방문연구원

2012년~2013년 강원대학교 기획부처장

2014년~현 재 강원대학교 IT대학 부학장

2005년~현 재 강원대학교 컴퓨터과학과 교수

관심분야: Data Mining & Knowledge Discovery, Stream Data,

Storage System, Database Applications, Mobile/

Wireless Communication Services & Systems