

Outlier Detection Method for Time Synchronization

Young Kyu Lee[†], Sung-hoon Yang, Ho Seong Lee, Jong Koo Lee, Joon Hyo Lee, Sang-wook Hwang

Center for Time and Frequency Metrology, Korea Research Institute of Standards and Science, 267 Gajeong-Ro, Yuseong-Gu, Daejeon 34113, Korea

ABSTRACT

In order to synchronize a remote system time to the reference time like Coordinated Universal Time (UTC), it is required to compare the time difference between the two clocks. The time comparison data may have some outliers and the time synchronization performance can be significantly degraded if the outliers are not removed. Therefore, it is required to employ an effective outlier detection algorithm for keeping high accurate system time. In this paper, an outlier detection method is presented for the time difference data of GNSS time transfer receivers. The time difference data between the system time and the GNSS usually have slopes because the remote system clock is under free running until synchronized to the reference clock time. For investigating the outlier detection performance of the proposed algorithm, simulations are performed by using the time difference data of a GNSS time transfer receiver corrected to a free running Cesium clock with intentionally inserted outliers. From the simulation, it is investigated that the proposed algorithm can effectively detect the inserted outliers while conventional methods such as modified Z-score and adjusted boxplot cannot. Furthermore, it is also observed that the synchronization performance can be degraded to more than 15% with 20 outliers compared to that of original data without outliers.

Keywords: outlier detection, synchronization, modified Z-score, GNSS, system time

1. INTRODUCTION

GNSS와 같은 위성 항법 시스템 또는 eLoran과 같은 지상 파 항법 시스템의 기준이 되는 시스템 기준시는 하나의 전역 기준 시각(일반적으로 세계협정시, UTC: Coordinated Universal Time)에 대해서 항법시스템을 구축할 때 요구되어지는 동기 정확도를 제공하기 위해 정해진 범위 내에서 동기를 유지해야 한다. 예를 들면, GPS 위성 항법시스템은 UTC(USNO)와 40 ns 이내에서 (Naval Oceanography Portal 2020), Galileo 시스템 타

임(GST: Galileo System Time)은 UTC에 50 ns 이내로 유지되며 (Bedrick et al. 2004, Hahn & Powers 2005), eLoran 시스템의 경우 UTC에 100 ns 이내로 동기 되도록 요구되어지고 있다 (LORIPP 2008). 이러한 항법 기준시를 정해진 범위 내에서 UTC와 동기를 시키기 위해서는 항법 기준시와 UTC 사이의 시각차 데이터를 생성해야 하고, 이를 토대로 원하는 동기 성능을 이루기 위해서 위상 및 주파수 오프셋을 추정하여 제어해 주게 된다. 따라서 만일 시각비교 데이터에 정상적인 데이터가 아닌 이상 데이터(outlier data)가 존재하게 되면 정확하지 않은 위상 및 주파수 오프셋을 추정하게 되어 동기 성능의 저하를 가져오게 된다. 이상 데이터는 관측된 다른 데이터들과 크게 다른 측정 값을 갖는 데이터를 의미하며, 측정 환경의 변동이나 측정시 발생하는 다양한 에러에 의해 기인한다 (Wikipedia 2020).

이상 데이터는 다양한 분야의 측정치에 존재하며, GPS를 이용한 위치 측정에 있어서 존재하는 이상 데이터를 검출하기 위한 최소 절대값(least absolute values) 및 일반화된 M-추정기(generalised M-estimators) 등과 같은 여러 가지 견실한 추정 방법에 대해서는 Knight & Wang (2009)에 자세히 나와 있고, GPS 수신기의 측정치에 존재하는 여러 기능적인 데이터를 세트화

Received Sep 22, 2020 Revised Oct 13, 2020 Accepted Oct 28, 2020

[†]Corresponding Author

E-mail: ykleeks@kriss.re.kr

Tel: +82-42-868-5569 Fax: +82-42-868-5569

Young Kyu Lee <https://orcid.org/0000-0002-6617-6689>

Sung-hoon Yang <https://orcid.org/0000-0002-6516-3210>

Ho Seong Lee <https://orcid.org/0000-0001-6416-2162>

Jong Koo Lee <https://orcid.org/0000-0001-6880-9189>

Joon Hyo Lee <https://orcid.org/0000-0001-7304-3624>

Sang-wook Hwang <https://orcid.org/0000-0002-7409-2409>

해서 H-model 알고리즘을 사용하여 기능 depth를 구하고 가장 depth가 낮은 데이터를 이상 데이터로 검출하는 방식이 (Bakar et al. 2006)에 기술되어 있다. 또한 이상 데이터의 검출은 시각 동기 데이터뿐만 아니라 여러 다양한 측정 분야에서도 사용되어지며, 이러한 분야 중 하나가 거대한 데이터 세트의 패턴을 찾는 데이터 채굴(mining)이며 제어 차트, 선형 회귀(regression) 및 Manhattan 거리 알고리즘 등이 (Ordonez et al. 2011)에 기술되어 있다. 이러한 방법들은 이상 데이터를 제거하기 위한 다양한 알고리즘을 설계하기 위해서 활용되어질 수 있지만, GPS 수신기가 제공하는 시각차 데이터에 존재하는 이상 데이터를 제거하기 위해 적용하기에는 적절하지 않다. 따라서 본 논문에서는 일반적으로 널리 사용되어지는 이상 데이터 검출 알고리즘을 기초로 하는 방안을 제안하였다.

측정 데이터 세트에 존재하는 이상 데이터를 판별하기 위한 방법으로는 데이터가 정규분포를 갖는다는 가정하에 평균과 표준편차를 활용하는 Z-score, Z-score의 평균과 분산이 극단적인 이상데이터에 크게 영향을 받는다는 단점을 보완하기 위한 수정(modified) Z-score, 데이터를 크기 순으로 정렬하여 상향 및 하향 경계치를 이용하는 boxplot, 비대칭성(non-symmetric)을 갖는 데이터를 견고하게 측정하기 위해 boxplot을 수정한 조정(adjusted) boxplot 등이 있다 (Seo 2006). 이러한 판별 방식들은 기술품을 가지지 않는 일반적인 데이터에 대해서는 상당히 유효한 방식이지만, 시각비교 데이터에 존재하는 주파수 오프셋에 의해 발생하는 기술품을 갖는 데이터에서는 유효하지 않는 상황이 발생한다. 따라서, 기술품이 존재하는 데이터에 대한 이상 데이터를 유효하게 검출할 수 있는 방안이 고정밀의 시각동기를 위해서 요구되어지며, 이를 위해 본 논문에서는 이웃하는 데이터 사이의 2-샘플 차이값을 사용하여 이상 데이터를 판별하는 방법에 대하여 기술하였다. 기존의 수정 Z-score와 조정 boxplot 및 제안하는 방법에 대해 모의실험을 통해 이상 데이터 검출 능력을 비교한 결과 기존의 방법들에서는 검출되지 않는 이상 데이터들도 제안하는 방식을 사용하면 유효하게 검출할 수 있다는 것을 관찰하였다.

본 논문의 제 2장에서는 이상 데이터를 검출하기 위한 기존의 Z-score, 수정 Z-score, boxplot 및 조정 Boxplot과 같은 전형적인 방법에 대해 설명한다. 제 3장에서는 실험실에서 운영 중인 GNSS 시각비교 데이터에 대한 기존의 수정 Z-score와 조정 boxplot에 대한 이상 데이터 검출 능력 및 제안하는 방식과의 비교 결과에 대해서 기술하고, 제 4장에서 결론을 맺는다.

2. OUTLIER DETECTION METHODS

2.1 Z-score

Z-score는 이상 데이터를 검출하기 위한 잘 알려진 방식 중 하나이고 다양한 분야에서 널리 사용되어지고 있으며, 수집된 데이터 세트가 정규 분포(normal distribution)를 갖고 있을 때를 기준으로 이상 데이터를 추정한다. n 개의 데이터 셋에 대한 Z-score는 식 (1)과 같이 정의 된다.

$$Z_{score}(i) = \frac{x_i - \bar{x}}{s}, \quad i = 1, 2, \dots, n. \quad (1)$$

여기서 $\bar{x}$ 는 평균값(mean)이고, $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ 는 샘플 표준편차(standard deviation)이다. 데이터 세트와 결정자의 판별 기준에 따라 달라지기는 하지만, 일반적으로 $|Z_{score}|$ 가 3보다 클 때 이상 데이터로 판정한다.

2.2 Modified Z-score

기존의 Z-score의 문제점은 이상 데이터를 판별할 때 사용되어지는 데이터의 평균과 표준편차가 극단적인 값을 갖는 이상 데이터에 큰 영향을 받는다는 것이다 (Leys et al. 2013). 따라서, 이러한 문제점을 보완하기 위해 제안된 것이 평균값 대신 중앙값(median)을 사용하고 표준편차 대신 Median of Absolute Deviations (MAD)를 사용하는 수정(modified) Z-score이며, 이에 대한 판별식은 식 (2)와 같다.

$$M_i = \frac{0.6745(x_i - \tilde{x})}{MAD}, \quad i = 1, 2, \dots, n. \quad (2)$$

여기서 $\tilde{x}$ 는 중앙값이고, $MAD = \text{median}(|x_i - \tilde{x}|)$ 이며, 분자의 0.6745는 큰 값의 n 에 대해 MAD의 평균값이 0.6745σ 이기 때문에 정규화(normalizing)를 위해 사용된다. 관측된 데이터 중 $|M_i| > D$ 이면 이상 데이터로 판정되고 일반적으로 $D = 3.5$ 가 널리 사용된다 (Iglewicz & Hoaglin 1993).

2.3. Boxplot

Boxplot (또는 Tukey's fences)은 하향 경계치와 상향 경계치를 사용하여 이 값을 벗어나는 데이터를 이상 데이터로 판별하며, 데이터를 연속적인 단일 변수로 표현하여 극단적인 데이터에 큰 영향을 받지 않는다. 이러한 boxplot은 중앙선 (일반적으로 중앙값; median)을 중심으로 하향 경계치 Q_1 과 상향 경계치 Q_3 를 구하고 펜스로 불리는 $k(Q_3 - Q_1)$ 값을 사용하여 이 값보다 큰 값을 이상 데이터로 검출한다. k 는 보통 1.5의 값이 사용되어지고, $Q_1 = x_f$ (f 번째 관측 데이터)이며, f 는 식 (3)과 같이 구한다 (Tukey 1997).

$$f = \frac{[(n+1)/2]+1}{2} \quad (3)$$

만일 f 가 정수가 아니면 Q_1 은 x_f 와 x_{f+1} 의 평균 값이며 $Q_3 = x_{n+1-f}$ 이다. Boxplot 방법을 사용하여 이상 데이터를 검출하는 순서는 다음과 같다.

- ① 첫 번째, 내부 4분위 IQR산출: $IQR = Q_3 - Q_1$;
- ② 두 번째, 내부 경계 구간 산출: $[Q_1 - 1.5IQR, Q_3 + 1.5IQR]$;
- ③ 세 번째, 외부 경계 구간 산출: $[Q_1 - 3IQR, Q_3 + 3IQR]$;
- ④ 네 번째, 내부 경계 구간과 외부 경계 구간에 있는 관측 데이터를 잠재적인 이상 데이터로 판별.

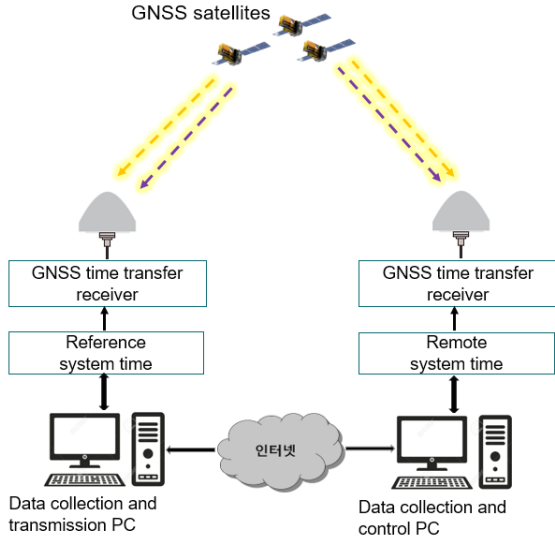


Fig. 1. Conceptual diagram of a time comparison by using GNSS timing receivers.

2.4 Adjusted Boxplot

관측된 데이터가 큰 비대칭성을 가지고 있을 때 효율적으로 이상 데이터를 제거하기 위한 방법 중 하나가 조정(adjusted) boxplot 이다 (Hubert & Vanderviere 2008). 조정 boxplot은 boxplot과 비슷하게 하향 경계치와 상향 경계치를 사용하여 이 경계치를 벗어난 데이터를 이상 데이터로 판정한다. 크기에 따라 정렬된 데이터 세트에 대한 Med-Couple (MC)는 식 (4)와 같이 정의되고

$$MC = \underset{x_i \neq x_j}{\text{median}}(h(x_i, x_j)) \quad (4)$$

여기서 함수 h 는 식 (5)와 같다.

$$h(x_i, x_j) = \frac{(x_j - \bar{x}) - (\bar{x} - x_i)}{x_j - x_i} \quad (5)$$

조정 boxplot의 하향 및 상향 구간은

$$[L, U] = \begin{cases} [Q_1 - 1.5e^{-3.5MC}(Q_3 - Q_1), Q_3 + 1.5e^{4MC}(Q_3 - Q_1)], & \text{if } MC \geq 0 \\ [Q_1 - 1.5e^{-4MC}(Q_3 - Q_1), Q_3 + 1.5e^{3.5MC}(Q_3 - Q_1)], & \text{if } MC \leq 0 \end{cases} \quad (6)$$

과 같이 주어지고 여기에서 L 은 하향 경계치이고, U 는 상향 경계치이며, Q_1 과 Q_3 는 boxplot에서 구한 것과 같다. MC의 값은 -1과 1 사이의 값을 갖고, 만일 MC가 0이면 데이터는 대칭적이고, 0보다 크면 우편향적이며 0보다 작으면 좌편향적이다.

3. PROPOSED METHOD FOR TIME COMPARISON DATA

GNSS 시각비교 수신기를 이용하여 원격지 클락을 기준국 클

락에 동기 시키기 위해 일반적으로 사용되어지는 개념도를 Fig. 1에 나타 내었다. Fig. 1에 나타난 것과 같이 원격지 및 기준국 모두에서 GNSS 위성 클락과 각각의 시스템 클락과의 시각차 데이터를 Common GPS GLONASS Time Transfer Standard (CGGTTS) 파일 형태로 생성한다. 원격지에서는 기준국에서 생성한 시각비교 데이터를 인터넷 등을 통해 수집하고 이를 사용하여 기준국 타임과 원격지 타임과의 시각차 데이터를 생성하며, 이렇게 생성된 시각차를 0으로 맞추기 위해서 원격지 클락을 제어한다. 이러한 시각비교 데이터의 특성은 원격지의 클락은 기준 시각과의 시각차 데이터를 생성하고 이를 사용하여 동기를 시켜주기 전까지 자유 구동(free running) 상태가 지속되기 때문에 원격지 클락의 주파수 오프셋에 의한 기울기가 발생한다는 것이다. 본 장에서는 먼저 시각차 데이터에 기울기가 존재했을 경우에 있어서 기존의 수정 Z-score와 조정 boxplot 알고리즘을 사용하였을 때의 이상 데이터 판별 능력에 대해 기술하였고, 다음으로 제안하는 방식의 이상 데이터 판별 능력 및 이상 데이터를 제거하지 않았을 때의 시각 동기 오차에 대해서 기술하였다.

3.1 Outlier Detection Capability of Conventional Methods

기울기를 갖는 데이터에 대한 이상 데이터 검출 능력을 관찰하기 위해서 자유 구동 중에 있는 세습 원자시계의 1 PPS 및 10 MHz 출력이 연결된 GNSS 시각비교 수신기를 이용하여 측정된 모의실험 데이터를 Fig. 2에 나타 내었다. Fig. 2에서는 편의를 위하여 초기 위상 오프셋을 0으로 조정했으며, 그림에서 볼 수 있는 것과 같이 자유구동하는 세습 원자시계가 기울기를 가지고 하루 동안에 80 ns 정도 변화하고 있고, 기울기가 제거 되었을 때에는 평균 0 근처에서 ± 15 ns 이내에서 변동하고 있음을 알 수 있다. 먼저, 기존의 수정 Z-score 방법을 이용하였을 때의 original 데이터에 대한 MAD, 중앙값 및 표준편차는 각각 19.9, 33.1 및 27.4 ns였고, 기울기가 제거된 residual 데이터에 대해서는 각각 3.2, 0.1 및 5.3 ns였으며, 두 경우 모두 검출된 이상 데이터는 없었다. 한편, 조정 Boxplot 방법을 이용하였을 때의 original 데이터에 대한 MC, 상향 경계 값 및 하향 경계 값은 각각 0.3, 226.7 및 3.9 ns였고, residual 데이터에 대해서는 각각 -0.1, 10.2 및 -15.5 ns였다. Original 데이터에 대해서는 하향 경계 값 3.9 ns보다 작은 0.0, 2.5, 1.5 ns와 같은 3개의 데이터가, residual 데이터에 대해서는 상향 경계 값 10.2 ns 보다 큰 13.6, 11.1, 10.9 ns와 같은 3개의 데이터가 각각 이상 데이터로 판별되었다. 이상의 결과를 종합해 볼 때, 주파수 오프셋에 의한 기울기가 존재하고, 또한 동기를 위한 steering에 의해 0에 가까운 데이터가 일반적으로 상시 존재하는 시각 비교 데이터에 대한 이상 데이터 검출 방식으로는 수정 Z-score가 조정 boxplot보다 더 적절한 알고리즘이라는 것을 관찰할 수 있었다.

3.2 Proposed Algorithm

Fig. 3에는 시각비교 데이터의 이상 데이터를 제거하기 위한 제안하는 방식에 대한 순서도를 나타내었다. Fig. 3에 나타난 것

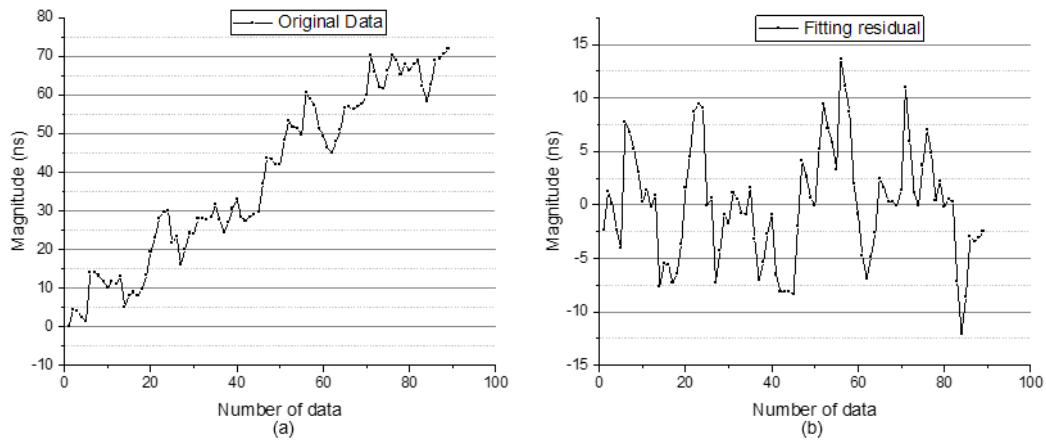


Fig. 2. Original and residual data for the simulation.

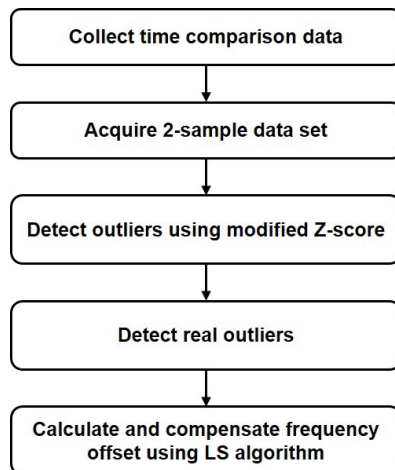


Fig. 3. Flowchart for detecting the outlier data of the proposed algorithm.

과 같이 시각비교 데이터를 수집한 이후에 먼저 $n-1$ 개의 2-샘플 데이터 세트를 획득한다. 이러한 데이터는 식 (7)을 이용하여 구한다.

$$y_i = |x_{i+1} - x_i|, i = 1, 2, \dots, n-1 \quad (7)$$

식 (7)에서 이웃하는 두 데이터 사이의 차이 값에 절대 값을 사용한 이유는 부호에 관계없이 정상적인 기울기 변화 이상의 큰 변화량을 갖는 데이터 세트를 구별하기 위함이다. 이러한 방법을 사용하게 되면 이상 데이터와 이전 데이터 및 이후 데이터와의 양쪽 차이 값을 이상 데이터로 판별하게 된다. 예를 들면, m 번째 데이터가 실제적인 이상 데이터라면 m 번째 데이터와 $m-1$ 번째 차이 값 및 $m+1$ 번째 데이터와 m 번째 데이터와의 차이 값에 대해 이상 데이터로 판별하게 된다. 따라서, 실제적인 이상 데이터를 검출하기 위한 추가적인 알고리즘이 필요하며, 2-샘플 데이터 세트의 수정 Z-score의 이상 데이터로 판별하는 문턱값을 MZ_{Th} 라 할 때, 이는 다음에 기술한 것과 같다. 이를 위해 이상 데이터가 연속적으로 인접하여 발생하지 않고 최소 2 샘플 이상 서로 떨어져 있다고 가정한다.

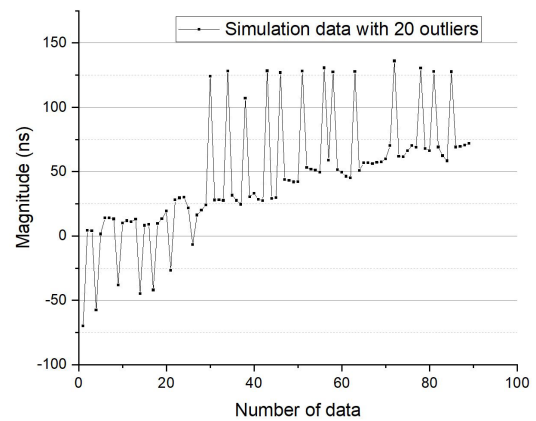


Fig. 4. Simulation data with intentionally inserted 20 outliers.

- ① 2-샘플 데이터의 첫 번째 데이터가 이상 데이터로 검출 되었을 경우:
 - 첫 번째 데이터(x_1)가 이상 데이터: $y_1 \geq MZ_{Th}$ 및 $y_2 < MZ_{Th}$ 일 때
 - 두 번째 데이터(x_2)가 이상 데이터: $y_1 \geq MZ_{Th}$ 및 $|x_3 - x_1| < MZ_{Th}$ 일 때
- ② 2-샘플 데이터의 첫 번째와 마지막 데이터가 아닌 중간 데이터가 이상 데이터로 검출 되었을 경우:
 - $i-1$ 번째 데이터(x_{i-1})가 이상 데이터: $y_{i-1} \geq MZ_{Th}$ 및 $y_i < MZ_{Th}$ 일 때
 - i 번째 데이터(x_i)가 이상 데이터: $y_i \geq MZ_{Th}$ 및 $|x_{i+2} - x_i| < MZ_{Th}$ 일 때
- ③ 2-샘플 데이터의 마지막 번째 데이터가 이상 데이터로 검출 되었을 경우:
 - 마지막 번째 데이터(x_n)가 이상 데이터: $y_{n-1} \geq MZ_{Th}$ 및 $y_{n-2} < MZ_{Th}$ 일 때
 - 마지막에서 두 번째 데이터(x_{n-1})가 이상 데이터: $y_{n-1} \geq MZ_{Th}$ 및 $|x_{n-2} - x_{n-1}| < MZ_{Th}$ 일 때

마지막으로, 위와 같은 절차를 거쳐 이상 데이터를 검출하고 이를 제거한 후에 최소제곱 (LMS: Least Mean Square) 알고리즘을

Table 1. Comparison of outlier detection capability between the conventional modified Z-score and the proposed method with outliers.

Parameter	Number of outliers					
	Modified Z-score (ns)			Proposed method (ns)		
	5	10	20	5	10	20
MAD	20.1	20.7	24.1	1.7	2.2	5.7
Median	33.1	37.0	43.9	2.5	2.8	6.0
Range for normal data	-71.2~	-70.4~	-81.2	-6.3~	-8.6~	-23.6
	137.4	144.3	168.9	11.3	14.2	35.6
Number of detected outliers	0	0	0	9	19	39

사용하여 위상 및 주파수 오프셋을 구한다.

이상 데이터가 존재할 경우에 제안하는 2-샘플 방법의 유효성을 기존의 방식과 비교하기 위해서 original 데이터에 5, 10 및 20 개의 이상 데이터를 임의로 넣은 데이터 중 20 개의 이상 데이터가 포함된 데이터를 Fig. 4에 나타 내었다. 이러한 이상 데이터의 크기는 수정 Z-score를 사용하였을 때, 정상적인 데이터로 검출되는 범위 내의 값을 사용하였다. 수정 Z-score를 사용했을 때 이상 데이터로 판별되는 데이터의 값은 식 (2)를 사용하여 $|M_i| > 3.5$ 인 값으로 구할 수 있으며, 식 (8)과 같다.

$$\begin{aligned} x_i &> MAD * 3.5/0.6745 + \tilde{x}, \text{ if } x_i \geq \tilde{x} \\ x_i &< \tilde{x} - MAD * 3.5/0.6745, \text{ if } x_i < \tilde{x} \end{aligned} \quad (8)$$

식 (8)을 이용하면 original 데이터에 대한 정상적인 데이터로 판별되는 범위는 $-70.2 < x_i \leq 136.4$ (ns)이다. 따라서 위의 범위 내의 값에 있도록 이상 데이터를 original 데이터에 추가해 주면 수정 Z-score를 사용해서는 이상 데이터로 검출되지 않는다.

Table 1에는 기존의 수정 Z-score와 제안하는 방식과의 이상 데이터 검출 능력에 대한 비교를 위해 정상 판정 범위 및 검출된 이상 데이터 수를 나타 내었다. Table 1에서 볼 수 있는 것처럼, 기울기가 존재하는 경우에 있어서 기존의 수정 Z-score 방식을 이용하였을 때에는 모든 경우에 있어서 이상 데이터를 하나도 검출하지 못했다. 반면에 제안하는 방식을 이용하면 모든 경우에 있어서 이상 데이터를 검출할 수 있었다.

Table 2에는 이상 데이터가 존재할 때 이를 제거해 주지 않고 제어를 해 주었다고 가정했을 때의 동기 성능 저하에 대해 살펴 보기 위해 original 데이터에 대한 위상 오차 값을 나타 내었다. Table 2를 보면 이상 데이터가 존재하지 않는 original 데이터는 선형 피팅에 의한 기울기와 절편에 따른 위상 변화량이 하루에 76.6 ns이지만 이상 데이터가 존재할 경우에는 이상 데이터의 개수가 커짐에 따라 변화량도 같이 커지며, 이상 데이터가 10개 이상일 때에는 10 ns 이상의 위상 오차를 발생시킬 수 있음을 관찰할 수 있다. 또한, 이상 데이터를 제거만하고 중앙값이나 평균 값 등과 같은 다른 값으로 대체하지 않았을 경우에는 이상 데이터 개수가 증가할수록 위상 오차가 커지는 하지만 그 값이 2 ns 정도 이내로 상당히 작은 위상 오차만을 가짐을 관찰할 수 있다. 따라서, 이상 데이터를 중앙값이나 예측된 기울기에 따른 다른 값을 대체하지 않고 사용하더라도 동기 성능에는 큰 차이가 없을 것으로 예상된다.

Table 2. Comparison of phase offset performance for the situations with outliers and without outliers.

Data types	Slope (ns/s)	Intercept (ns)	Phase variation (ns)	Phase offset (ns)
Original	8.46x10 ⁻⁴	2.0	76.6	0
Outlier exist	5 10.10x10 ⁻⁴	-4.8	82.5	5.9
	10 11.30x10 ⁻⁴	-8.5	89.2	12.6
	20 10.90x10 ⁻⁴	0.1	94.3	17.7
Outlier removed	5 8.51x10 ⁻⁴	1.8	75.3	-1.3
	10 8.44x10 ⁻⁴	2.2	75.1	-1.5
	20 8.34x10 ⁻⁴	2.7	74.8	-1.8

4. CONCLUSIONS

본 논문에서는 원격지의 시스템 클락을 동기 기준이 되는 기준 시(예, 세계협정시) 클락에 동기 시키기 위해 사용되어지는 시각비교 데이터에 존재하는 이상 데이터를 효과적으로 제거하는 방법에 대하여 기술하였다. 이러한 시각비교 데이터에는 일반적인 데이터와는 다르게 GNSS 시각비교 수신기에 기준 신호를 제공하는 원자시계(또는 양상블 시스템 클락)에 존재하는 주파수 오프셋에 의해 기울기가 존재하게 된다. 또한 원격지 클락을 기준시에 동기 시키는 과정에서 주파수 오프셋을 조절해 주어야 하기 때문에 시각비교 데이터에 있어서 기울기는 상시 존재하게 된다. 따라서, 기울기에 의한 영향을 효과적으로 감소시켜 주기 위한 방법이 필요하며, 이를 위해 인접한 두 데이터 샘플의 차이 값을 활용하여 이러한 차이 값들의 2-샘플 데이터 세트를 구하였다. 이렇게 구해진 2-샘플 데이터 세트는 정상적인 데이터에 대해서는 기울기와 기울기가 제거된 데이터 사이의 변동폭 이내의 값을 갖는 반면 이상 데이터에 대해서는 정상 범위를 벗어나는 값을 갖게 되어 기존의 수정 Z-score를 사용하여 이상 데이터를 검출할 수 있게 된다.

또한, 이상 데이터를 검출하기 위해 일반적으로 널리 사용되어지고 있는 수정 Z-score나 조정 boxplot과 같은 방법의 유효성을 관찰하기 위해 실험실에서 실제 운영 중인 GNSS 시각비교 데이터에 임의의 크기를 갖는 이상 데이터를 추가하여 모의실험을 수행하였다. 모의실험을 통해 다음과 같은 사항을 관찰하였다. 첫째, 조정 boxplot의 경우 이상 데이터가 존재하지 않는 original 데이터에 대해서도 이상 데이터로 판별하는 경향이 있었다. 이러한 이유는 수정 boxplot은 데이터를 크기 순으로 배치한 후 하향 경계치 또는 상향 경계치를 벗어나는 데이터를 이상 데이터로 판별하는데 시각비교 데이터에는 0에 가까운 아주 작은 값들이 존재하게 되기 때문에 이를 이상 데이터로 판별하게 되는 것이다. 따라서, 조정 boxplot은 시각비교 데이터에 존재하는 이상 데이터를 판별하는 것으로는 적절하지 않다는 것을 알 수 있다. 둘째, 수정 Z-score의 경우에 있어서 original 데이터에 대한 판별에 있어서는 이상 데이터가 검출되지 않았지만, 이상 데이터의 판별에 사용되는 문턱값 이내의 크기로 추가된 임의의 이상 데이터들은 검출하지 못하였다. 따라서, 기울기가 있는 시각비교 데이터에 적용하는 데에는 적절하지 않다는 것을 알 수 있었다. 셋째, 제안하는 2-샘플 데이터 세트를 구한 후 수정 Z-score를 사용하면 기존의 방식들에서는 검출하지 못하는 이상 데이터들을 효과적으로 검출할 수 있음을 확인하였다. 마지막으로, CGGTTS에서 제공하는 하루 동안의 측정 데이터 88개의 22% 정도인 20개의 이

상 데이터가 제거되지 않았을 때에는 대략 23% 정도의 위상 오차를 발생시키는 것을 관찰할 수 있었다. 물론 모의실험에 사용된 이상 데이터의 크기와 위치에 따라서 위상 오차의 크기는 달라지겠지만 이상 데이터가 동기 성능에 큰 영향을 끼칠 수 있다는 것은 명확하게 확인할 수 있었다.

본 논문에서 제안하는 이상 데이터 검출 방식을 사용하면 기울기가 존재하는 시각비교 데이터에서 효과적으로 이상 데이터를 제거할 수 있기 때문에, 기준시(세계협정시(UTC) 또는 한국표준시(UTC(KRIS))에 고정밀의 시각동기를 이루어야만 하는 위성/지상파 항법시스템의 기준시 생성 시스템에서 주요하게 활용 가능할 것으로 예상된다.

AUTHOR CONTRIBUTIONS

Methodology and software, Y.K. Lee; data acquisition, S.-h. Yang, H.S. Lee; measurement system, J.K. Lee, J.H. Lee, S.-w. Hwang.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Bakar, Z. A., Mohamad, R., Ahmad, A. & Deris, M. M. 2006, A Comparative Study for Outlier Detection Techniques in Data Mining, IEEE Conference on Cybernetics and Intelligent Systems, available from: <https://www.researchgate.net/publication/224573269>
- Bedrick, S., Bauch, A., Moudrak, A. & Schafer, W. 2004, Design of the Precise Time Facility for Galileo, 36th Annual Precise Time and Time Interval (PTTI) Meeting, pp.293-306.
- Hahn, J. H. & Powers, E. 2005, Implementation of the GPS to Galileo time offset (GGTO), Joint IEEE International Frequency Symposium and Precise Time and Time Interval (PTTI) Systems and Applications Meeting, 29-31 Aug 2005, Vancouver, BC, Canada, pp.33-37. <https://doi.org/10.1109/FREQ.2005.1573899>
- Hubert, M. & Vandervieren, E. 2008, An Adjusted Boxplot for Skewed Distributions, Computational Statistics & Data Analysis, 52, 5186-5201. <https://doi.org/10.1016/j.csda.2007.11.008>
- Iglewicz, B. & Hoaglin, D. C., 1993, How to Detect and Handle Outliers, ASQC Basic References in Quality Control, vol.16 (Milwaukee: ASQC Quality Press)
- Knight, N. L. & Wang, J. 2009, A Comparison of Outlier Detection Procedures and Robust Estimation Methods in GPS Positioning, Journal of Navigation, available from: <https://www.researchgate.net/publication/231786405>
- Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. 2013, Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, Journal of Experimental Social Psychology, 49, 764-766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- LORIPP/LORAPP, 2008, Specification of the eLoran System, RTCM Paper 196-2008-SC127-036
- Naval Oceanography Portal: USNO GPS Time Transfer [Internet], cited 2020 Aug. 25, available from: <https://www.usno.navy.mil/USNO/time/gps/usno-gps-time-transfer>
- Ordonez, C., Martinez, J., Rodriguea-Perez, J. R. & Reyes, A. 2011, Detection of Outliers in GPS measurements by Using Functional-Data Analysis, Journal of Surveying Engineering, 137, 150-155
- Seo, S. 2006, A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets, Master Thesis, Pittsburgh University, Pennsylvania, pp.1-39.
- Tukey, J. W. 1977, Exploratory Data Analysis (London: Addison Wesley).
- Wikipedia 2020, Outlier [Internet], cited 2020 Aug. 25, available from: <https://en.wikipedia.org/wiki/Outlier>



Young Kyu Lee received the Ph. D. degree in information and communications engineering from Gwangju Institute of Science and Technology, Gwangju, Korea, in 2002. In 2002, he joined the Time and Frequency group of the Korea Research Institute of Standards and Science (KRISS), Korea. He has been engaged in the research and development work field in precise time comparison via satellites and time distribution using ground-waves, since he participated in the satellite applications laboratory.



Sung-hoon Yang received the Ph. D. degree in electronic engineering from Chungnam National University, Daejeon, Korea, in 2012. He has been working at the Time and Frequency Group, Korea Research Institute of Standards and Science (KRISS), Daejeon, since 1984. He has worked on the development of a primary atomic clock and standard time keeping. His current activity is research on time transfer using satellites.



Ho Seong Lee received his Ph.D. degree in Physics from the Korea Advanced Institute of Science and Technology (KAIST) in 1986. He joined the Time & Frequency Group of the Korea Research Institute of Standards and Science (KRISS) in 1986. He worked for the development of cesium atomic clocks for about 20 years. His current interest includes the generation of timescales for the Korea Standard Time and the Korea Positioning System Time.



Jong Koo Lee graduated from the Department of Statistics and Data Science at Cheongju University, in 1989. He has been working at the Time and Frequency Group, Korea Research Institute of Standards and Science (KRISS), Daejeon, since 1990. His interesting fields are calibration of timing systems and statistical estimation.



Joon Hyo Lee received the B.S degree in electrical and electronic engineering and PH. D. degree in integrated technology at Yonsei University, Incheon, South Korea in 2019. He is currently a researcher with the Korea Research Institute of Standards and Science (KRISS), Daejeon, South Korea. He is now working at the Time and Frequency Group in KRISS, taking part in researches related to precise timing systems.



Sang-wook Hwang received the Ph. D. degrees in Electronics Engineering from Chungnam National University, Daejeon, Korea, in 2019. He has been working at the Time and Frequency Group, Korea Research Institute of Standards and Science (KRISS), Daejeon, since 2015. His current activity is research on time transfer using satellites.

