

# Wavelet 과 PCA 기법을 이용한 효율적 데이터 전송기법 개발에 관한 연구

## Study on the Development of effective data transmission Scheme based on Wavelet and PCA

육의수\* · 한윤종\* · 김성호\*

\* 군산대학교 전자정보공학부

Yui-Su Youk, Yoon-Jong Han and Sung-Ho Kim

School of Electronics and Information Engineering, College of Engineering

KunSan National University

E-mail : sixofnum@hanmail.net

### 요 약

최근 인터넷 및 무선 통신기술의 광범위한 보급으로 인해 현장 계측 데이터 등과 같은 중요 데이터를 인터넷을 통해 실시간으로 수신가능케 하는 다양한 형태의 웹 기반 원격 모니터링 시스템이 설계되고 있다. 이러한 웹 모니터링 시스템은 기본적으로 짧은 주기마다 측정된 데이터를 원격의 서버로 전송하는 것이 바람직하나 과도한 통신비 문제로 인해 효율적인 시스템 운영이 어렵다는 문제점을 갖는다. 따라서 본 연구에서는 측정데이터의 변화를 효율적으로 검출할 수 있는 PCA(Principle Component Analysis) 기법과 데이터 압축에 탁월한 특성을 갖는 wavelet 기법을 융합한 새로운 형태의 웹 기반 원격모니터링용 데이터 전송기법을 제안하고 실제 데이터에 적용하여 봄으로써 제안된 기법의 유용성을 확인하고자 한다.

Key Words : PCA(Principle Component Analysis), Wavelet, 원격 모니터링 시스템

### 1. 서론

최근 광범위 하게 보급된 인터넷 및 무선통신 기술은 효율적인 결합을 통해 다양한 산업분야에서 사용되고 있다. 특히 계통 관리자가 시간 및 장소에 구애받지 않고 현장의 운영상황을 신속히 모니터링 할 수 있도록 하는 인터넷 기반의 원격 모니터링 시스템은 무인화된 계측 현장 등에 효율적으로 도입되고 있다. 이러한 웹 모니터링 시스템은 크게 현장에 설치되어 다양한 정보를 수집하는 계측 스테이션과 이들 데이터를 원격의 서버로 전송하는 모뎀, 그리고 이들 데이터를 수신하여 다수의 계통관리자에게 정보를 제공하는 서버 시스템으로 구성된다.

일반적으로 현재 운영되고 있는 모니터링 시스템은 규정된 시간마다 무선 통신을 이용하여 측정된 데이터를 원격의 서버로 전송하는 방식으로 운영되고 있다. 하지만 이러한 운영은 과도한 통신비를 초래하게 되어 시스템 운용에 많은 문제를 일으키게 된다. 그러나 일반적으로 계측 현장에서 측정되는 센서 데이터들은 상호간에 선형 또는 비선형의

결합 관계를 갖게 되며 따라서 매 시간마다의 규칙적인 데이터 전송은 전송 효율을 저하하게 되는 요인이 되고 있다. 현재까지 많은 연구자들은 이러한 전송 효율의 문제 해결을 위해 다양한 압축 기법에 대한 연구를 진행시켜 왔다. 이중 Wavelet 압축 기법은 데이터의 압축 및 복원에 탁월한 성능을 갖는 것으로 알려져 있다[1-3]. Wavelet은 데이터의 압축을 위해 측정 데이터를 관련 벡터들로 분해하고 이를 전송함과 동시에 전송된 데이터는 재구성됨으로써 원래의 데이터를 복원할 수 있게 된다는 특징을 갖는다. Wavelet 중에서도 다중해상도기법으로 여러 단계의 레벨로 분해 및 재구성이 가능한 Daubechies 방법이 데이터 압축에 효율적으로 사용되고 있다.

PCA는 선형 시스템에 적용되는 기법으로 다변수 시스템의 중요한 변량의 방향을 결정하는 통계학적 기법이다. 이 기법은 센서의 이상 데이터 검출을 위한 유용한 방법으로 알려져 있다.

따라서 본 연구에서는 원격 모니터링 시스템의 전송 효율 직감대책으로 정상시의 계측 데이터와 이상이 발생한 경우의 데이터를 효율적으로 판별한

수 있게 하는 PCA 기법과 전송 데이터의 효율적 압축을 가능케 하는 wavelet 기법을 융합한 새로운 형태의 전송 알고리즘을 제안하고자 하며 실제 계통에의 적용을 통해 제안된 기법의 유용성을 확인하고자 한다. 본 논문의 구성은 다음과 같다. 제2장에서는 PCA의 기본 개념에 대하여 기술하며 제3장에서는 Wavelet 변환에 대해 기술하며 제4장에서는 본 연구에서 제안된 압축알고리즘인 Wavelet과 PCA기법의 적용에 대해 기술하고 5장에서는 결론을 기술하고자 한다.

## 2. 주성분 분석(Principle Component Analysis)

주성분 분석은 변수들의 선형 결합을 통해 변수들 간에 존재하는 중요한 변량의 방향을 결정하는 통계학적 모델링 기법으로 상관관계가 있는 변수들 사이의 복잡한 관계를 단순화하기 위하여 사용된다. 즉, 상관관계가 있는 변수들을 결합하여 서로 독립적인 새로운 저차의 새로운 변수들을 생성하는 것으로 생각할 수 있으며 이때 새로이 생성된 변수를 주성분이라 부른다. PCA의 기본 개념에 대해 설명하면 다음과 같다.

$n$  개의 변수들에 대해  $m$ -회 측정으로 얻어진 데이터 행렬을  $X(m \times n)$ 라고 정의하고  $X$ 는 0 인 평균과 단위 분산을 갖는다고 가정한다. 이 경우  $X$ 의 공분산 행렬은 다음과 같이 정의된다.

$$cov(X) = \frac{X^T X}{m-1} \quad (1)$$

행렬  $X$ 는 다음과 같이 분해할 수 있으며

$$X = \mathcal{X} + E \quad (2)$$

여기서  $\mathcal{X}$ 는 다음과 같다.

$$\mathcal{X} = \sum_{i=1}^r t_i p_i \quad (r \leq n) \quad (3)$$

또한 식(2)에서  $E$ 는 PCA 모델의 예측 편차를 나타낸다. 식(3)의 scores로 불리는 벡터  $t_i$ 는 프로세스에 존재하는 잡음에 대한 정보를 포함하며 벡터  $p_i$ 는 변수들이 어떻게 관계되어 있는지에 대한 정보를 나타낸다. 벡터  $p_i$ 는 다음과 같이  $X$ 의 공분산 행렬의 고유벡터이기도 하다.

$$cov(X) P_i = \lambda_i P_i \quad (4)$$

식(3)에서 사용되는  $r$ -개의 고유벡터( $p_i$ )는 공분산 행렬에 대한  $n$ -개 고유값 중 가장 큰 고유값에서부터  $r$ -개까지의 고유값에 해당하는 고유벡터

를 의미한다. 또한 각 고유벡터에 대한 가중치 벡터  $t_i$ 는 다음과 같이 구할 수 있다.

$$t_i = X p_i \quad (5)$$

식(3)으로 표현되는 PCA 모델을 이용한다면 새로운 측정벡터  $x(1 \times n)$ 에 대한 PCA 모델링 편차  $e$ 는 다음과 같이 구할 수 있으며

$$e = (x - \hat{x}) \quad (6)$$

새로운 측정 벡터  $x$ 에 대한 잔차  $R$ 은 다음과 같이 구할 수 있다.

$$R = e e^T \quad (7)$$

상기 식들을 이용한다면 프로세스의 정상동작에 대한 PCA 모델을 구하는 것이 가능하다. 따라서 정상시와는 다른 계측 데이터가 얻어진 경우 식(7)로 표현되는 잔차  $R$ 의 고찰을 통해 정상시와 다른 계측 데이터의 발생을 효율적으로 검출할 수 있게 된다. 따라서 식(7)에 대한 임계치 검사를 통해 효율적인 이상 데이터 발생의 검출이 가능하게 된다.

## 3. Wavelet 변환

Wavelet 변환은 주어진 신호를 스케일이 다른 각각의 성분들로 분해하는 선형 연산 과정이다. Wavelet 변환은 주어진 입력신호에 대해 Wavelet 함수와 스케일 함수를 각각 곱하여 고주파와 저주파로 신호의 주파수 대역을 나누어 원하는 주파수 대역에서의 변환된 신호로 분석할 수 있기 때문에 시간-주파수 국부성이 매우 우수한 특징을 갖는다. Wavelet 변환은 주어진 신호를 식(8)의 조건을 만족하는 기저함수(basis function)의 선형조합으로 표현하는 과정이다.

$$\int_{-\infty}^{\infty} \frac{|\overline{\Psi}(w)|^2}{|w|} < \infty \quad (8)$$

Wavelet 변환에 사용되는 기저함수로서는 스케일링 함수와 mother Wavelet이 있으며 이는 식(9)로 표현된다.

$$\Psi(x)_{a,d} = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-d}{a}\right) \quad (9)$$

식(9)에서  $a$ 는 scaling요소이고  $d$ 는 천이 요소이며  $\frac{1}{\sqrt{a}}$ 는 정규화 인자이다.

Wavelet 변환은 연속 Wavelet 변환과 이산 Wavelet 변환으로 분류된다. 신호  $s(t)$ 에 대한 연속시간 영역에서 Wavelet 변환(CWT: Continuous

Wavelet Transform)과 이산 Wavelet 변환은 각각 다음과 같다.

$$CWT(a, d) = \frac{1}{\sqrt{a}} \int s(t) \Psi^* \left( \frac{t-d}{a} \right) dt = \langle s(t), \Psi_{a,d}(t) \rangle \quad (10)$$

식 (9)에서 스케일링 파라미터와 천이 파라미터가 다음과 같이 표현될 경우 이산 Wavelet 변환은 다음과 같이 얻어진다.

$$a = 2^m, d = n2^m \quad m, n \in \text{정수} \quad (11)$$

$$DWT(m, n) = 2^{-\frac{m}{2}} \sum_k s(k) \Psi \left( 2^{-\frac{m}{2}} k - n \right) \quad (12)$$

mother-Wavelet의 형태는 Morlet, Haar, Shannon, Meyer, Daubechies 등 만들어 낸 사람들에 따라서 다양하며 각각의 mother Wavelet이 가지고 있는 신호 분해의 특성과 장점에 따라서 응용되는 분야는 다를 수 있다. 측정된 센서의 데이터들은 그 크기가 작고 짧은 지연특성과 빠른 감쇠, 그리고 진동하는 형태의 신호이므로 신호의 형태를 고려하여 'Daubechies 3'를 적용하였다. 그림 1과 같이 데이터 신호를 여러 형태의 고역필터 성분들로 나누기 위해서 Wavelet 필터 뱅크를 사용하여 3단계의 주파수 대역별(Fs/2, Fs/4, Fs/8)로 나누어서 분할처리할 수 있다. 즉 이산 Wavelet 변환을 이용하여 데이터 신호를 근사 공간의 신호(C<sub>n1</sub>, C<sub>n2</sub>, C<sub>n3</sub>)와 세부공간의 신호(d<sub>n1</sub>, d<sub>n2</sub>, d<sub>n3</sub>)로 분할할 수 있다. 근사공간의 신호는 high-scale로 저주파 성분이며 신호의 고유한 특성으로 크기와 해상도를 변화시킬 때 각 조건에서의 파형의 변화된 상태만 나타내며 세부 공간의 신호는 low-scale로 고주파 성분이며 신호의 미세한 특성을 나타내는 중요한 정보를 제공하여 준다. 분해된 신호는 그림 2와 같이 재구성될 수 있다.

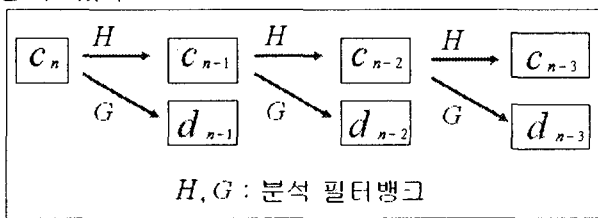


그림 1 웨이블릿 분해를 위한 필터뱅크 구조  
Fig. 1. Filter bank structure for Wavelet analysis

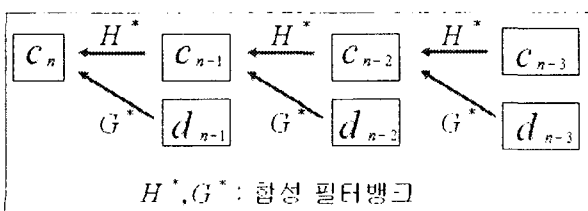


그림 2 웨이블릿 합성을 위한 필터뱅크 구조  
Fig. 2. Filter bank structure for Wavelet synthesis

#### 4. 웹 모니터링 시스템에의 적용

전 절에서 고찰한 PCA 및 wavelet 기법을 이용한 효율적 데이터 전송기법을 적용할 대상 시스템은 다음과 같다.

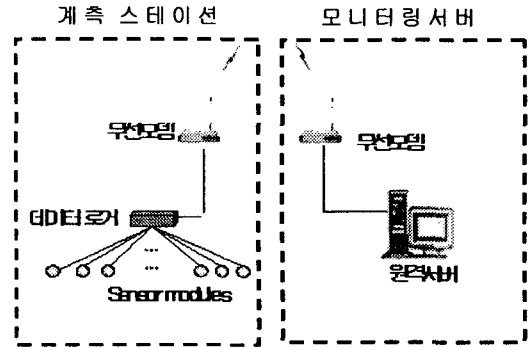


그림 3. 원격 모니터링 시스템 구성도  
Fig. 3. The structure of remote monitoring system

그림 3의 좌측은 계측현장에 설치되는 계측 스테이션을 나타내며 이의 전체 동작은 다음과 같다.

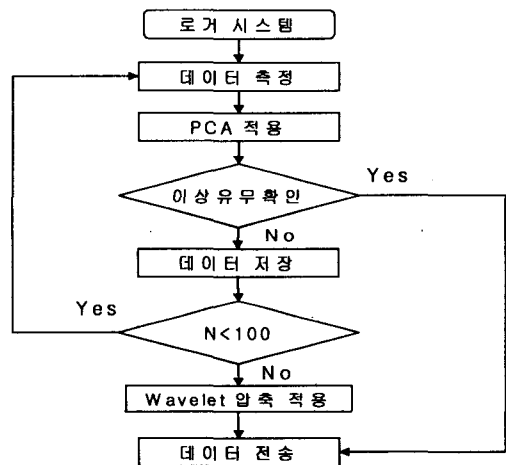


그림 4. 계측 스테이션의 동작

Fig. 4. Flowchart for measurement station

위의 그림4에서 보듯이 측정 데이터에 PCA기법을 적용하여 데이터의 이상 유무를 확인하고 측정 데이터에 이상이 없을 경우에는 일정한 측정간격으로 데이터를 저장해 두었다가 압축하여 전송하게 된다. 만약 이상이 검출된다면 wavelet을 이용한 압축을 수행하지 않고 직접 실시간으로 원격의 서버로 측정 데이터를 전송하게 된다.

#### 4.1 사면 계측 데이터에의 적용

제안된 기법의 유용성 확인을 위해 경사면에 설치된 35개의 센서 신호를 사용하고자 한다. 35개의 센서중 15번째부터 20번째의 센서 출력을 나타내면 그림 5와 같다.

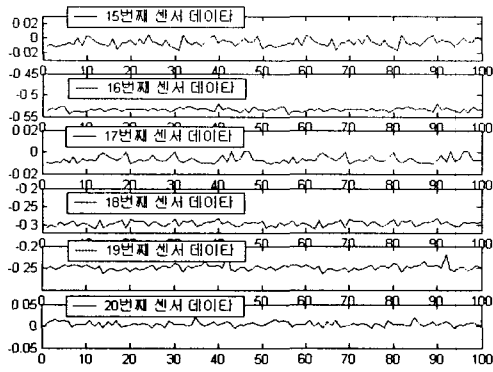


그림 5. 정상상태의 사면 계측 데이터  
Fig. 5. Normal slope data

만일 18번째 센서 출력에 55번째 스텝에서 램프 형태의 바이어스를 인가한다면 식 (7)로부터 얻어지는 잔차는 다음과 같다.

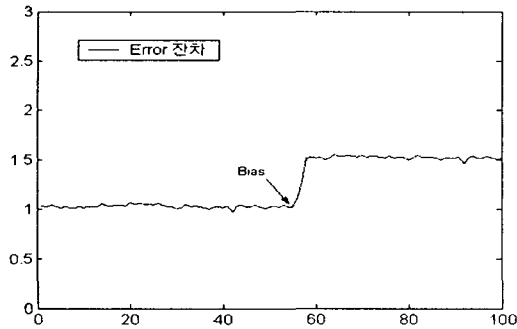


그림 6. PCA에 의한 이상상태의 검출  
Fig. 6. Detection of abnormal data by PCA

그림 6으로부터 알 수 있듯이 센서 데이터에 발생한 바이어스 형태의 이상데이터는 식 (7)의 잔차에 대한 임계치 검출을 통해 쉽게 검출할 수 있게 된다. 이와 같이 임계치 검사에 의해 이상 데이터가 검출된 경우에는 Wavelet에 의한 데이터 압축을 수행하지 않고 규정된 시간 간격으로 원격의 서버로 측정데이터를 실시간으로 전송하게 된다.

그러나 식 (7)의 잔차가 임계치를 넘지 않을 경우 측정데이터는 규정된 횟수만큼 메모리에 저장되었다가 wavelet을 통해 데이터 압축이 수행된다. 그림 7은 저장된 데이터의 근사계수와 상세 계수를 나타낸다.

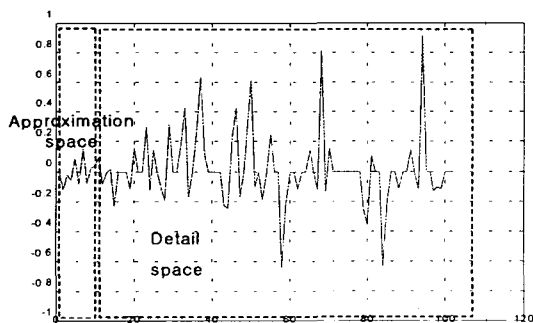


그림 7. 분해된 근사 계수와 상세계수  
Fig. 7. Approximation coefficient and detail coefficient

## 4.2 원격 서버에서의 압축 데이터의 복원

원격 서버에서는 계측 스테이션으로부터 전송된 압축 데이터를 복원하여 원래의 데이터를 얻게 된다. 그림 8은 계측 스테이션으로부터 압축되어 전송된 데이터를 복원한 것과 실제 측정 데이터를 나타낸다.

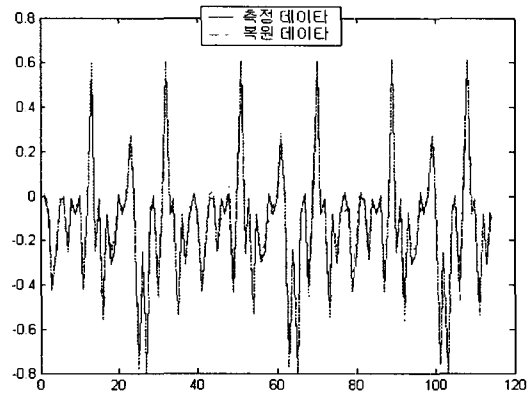


그림 8. 측정 데이터와 복원된 데이터  
Fig. 8. Measure data and reconstructed data

## 5. 결 론

본 논문에서는 원격 모니터링 시스템의 데이터 전송 효율 향상을 위해 PCA기법 및 Wavelet 기법을 융합한 새로운 형태의 전송기법을 제안하였으며 사면 계측 데이터에의 적용을 통해 유용성을 확인하였다. 시뮬레이션 결과를 통하여 PCA 기법에 의한 효율적인 이상 데이터의 검출이 가능함과 wavelet를 이용한 효율적인 데이터의 압축 및 복원이 가능함을 알 수 있었다.

본 연구는 산업자원부의 지원에 의하여 기초전력 연구원(R-2003-B-274) 주관으로 수행된 과제임

## 참 고 문 헌

- [1] Wavelet and Wavelet Packet Compression of electrocardiograms, Michael L. Hilton
- [2] Discrete Wavelet Transforms: Theory and Implementation, Tim Edwards, 1992
- [3] Parallel Wavelet-Based Compression Of Two-Dimensional Data, Marica Lucka, 2000
- [4] Dunia, R., S.J. Qin, "Identification of faulty sensors using principle component analysis", AIChE J., Vol.42, No.10, pp. 2797-2812, 1996
- [5] Jolliffe, I.T., Principal Component Analysis, New York Springer-Verlag.