

센서 네트워크를 위한 지능형 데이터 유효화 기법의 개발

Development of Intelligent Data Validation Scheme for Sensor Network

육 의 수*, 김 성 호

(Yui-Su Youk and Sung-Ho Kim)

Abstract : Wireless Sensor Network(WSNs) consists of small sensor nodes with sensing, computation, and wireless communication capabilities. The large number of sensor nodes in a WSN means that there will often be some nodes which give erroneous sensor data owing to several reasons such as power shortage and transmission error. Generally, these sensor data are gathered by a sink node to monitor and diagnose the current environment. Therefore, this can make it difficult to get an effective monitoring and diagnosis. In this paper, to overcome the aforementioned problems, intelligent sensor data validation method based on PCA(Principle Component Analysis) is utilized. Furthermore, a practical implementation using embedded system is given to show the feasibility of the proposed scheme.

Keywords : Wireless Sensor Network(WSN), Principle Component Analysis(PCA), embedded system, diagnosis

I 서론

인터넷의 급속한 보급으로 정보기술(IT)은 기업의 업무, 생산, 제조에서부터 국가의 공공, 교육에 이르기까지 광범위하게 적용되고 있다. 특히 컴퓨터와 인간과의 관계에서 이러한 정보 기술은 제 3의 물결, 즉 유비쿼터스 컴퓨팅으로 진화하고 있다. 초소형의 컴퓨터 디바이스를 사물에 내장하여 물리공간과 네트워크 기반의 가상공간을 융합한 새로운 유비쿼터스 컴퓨팅 세계는 인간의 생활을 보다 안전하고, 생산적이며, 풍요롭게 변화시키고 있다. 이러한 유비쿼터스 사회를 구현하기 위한 한 방향으로 최근 센서네트워크가 각광을 받기 시작하고 있다.

센서네트워크의 구축에는 계측, 연산 및 무선 통신기능을 갖춘 저가격, 저전력의 센서 노드들이 사용되고 있다. 이러한 센서 노드들은 사람이 접근하기 어려운 작업공간에 살포되며 노드들 자체적으로 네트워크를 형성함으로써 측정된 정보를 형성된 네트워크를 통해 인터넷과 같은 기간망으로 전달하게 된다. 이로 인해 사용자들은 원격지에서 자연환경의 변화를 효과적으로 관측할 수 있게 되며 이로 인해 센서 네트워크는 환경, 생태 모니터링, 지진 감시 및 군사용 등에 폭넓게 도입되고 있다[1].

광범위한 측정 현장에 설치되는 센서 노드는 여러 가지 요인(노드의 배터리 고갈, 사고로 인한 보드의 파손, 센서의 노화 등)으로 인해 측정 데이터의 신뢰도를 저하시킬 수 있으며 이런 측정 데이터에 포함된 불확실성 및 오류는 모니터링 시스템의 성능을 저하시키는 요인이 될 수 있다.

일반적으로 신뢰도가 높은 분석 결과를 얻기 위해서는 이상 데이터의 신속한 검출 및 이의 효율적 보상을 가능케 하는 자동화된 센서 데이터 유효화 기법의 개발이 요구된

다[2]. 최근 고차의 다변수 데이터들간에 존재하는 중요한 변량을 저차의 주성분들로 표현할 수 있게 하는 PCA(Principle Component Analysis)가 제안된 바 있다. 특히, PCA는 측정 데이터들로부터 얻어지는 공분산 행렬로부터 측정 데이터에 대한 PCA 모델을 간단히 구할 수 있으며 PCA 모델과 실제 측정 데이터 간의 잔차를 고찰함으로써 데이터에 포함된 異常을 효과적으로 검출할 수 있다는 특징을 갖는다[3-6].

이에 본 연구에서는 Multi-hop 기반의 무선 센서 네트워크 상황하에서 임베디드 기반의 sink 노드에서 수집된 데이터에 포함된 각종 불확실성을 효과적으로 검출할 수 있는 PCA 기반의 데이터 유효화 알고리즘을 제안하고 실제 임베디드 시스템에 적용하여 봄으로써 제안된 기법의 유용성을 확인하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 센서 네트워크의 구성요소에 대해 설명하며, 3장에서는 데이터의 불확실성을 검출하기 위한 PCA알고리즘에 대해 설명하고, 4장에서는 센서 네트워크의 유효화 기법에 관한 전체 구성 및 원리에 대해 설명하며, 5장에서는 실제 구현된 시스템에서의 동작 및 성능 검토를 기술하고 마지막으로 6장에서 결론을 기술한다.

II 센서 네트워크 구성 요소

센서네트워크란 인간 친화적인 미래 컴퓨팅 패러다임인 "유비쿼터스 컴퓨팅"을 실현 가능케 하는 기반 네트워킹 기술로서 광범위한 지역에 분산·설치되는 초소형, 저전력의 센서 노드들로 구성되며 이들 각각에 대한 설명은 다음과 같다.

1. 센서 노드

본 연구에서 사용된 센서 노드는 한백전자에서 개발된 2.4GHz 대역 통신을 사용하는 ZigbeX를 사용하였으며, 이러한 ZigbeX의 메인보드는 그림 1과 같다.

ZigbeX는 Atmel사의 저전력 IC인 Atmega128L MCU와

* 책임저자(Corresponding Author)

논문접수 : 2006. 12. 25., 채택확정 : 2007. 3. 1.

육의수, 김성호 : 군산대학교 전자정보공학부

(sixofnum@kunsan.ac.kr/shkim@kunsan.ac.kr)

※ 이 논문은 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. R01-2005-000-11088-0).

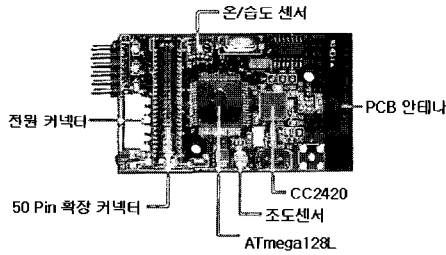


그림 1. ZigBee 메인보드.
Fig. 1. ZigBee main board.

Chipcon사의 CC2420 을 사용한 sensor network module이다. ZigbeX 센서 노드에는 온/습도 센서와 조도센서가 내장되어 있어 주변 환경 정보를 확인할 수 있으며 50Pin 커넥터를 이용하여 외부모듈들과 통신할 수 있도록 설계되었다. 전원은 AA battery를 이용하거나 기타 외부 전원을 선택적으로 사용하도록 설계되었으며 전원의 효율적인 사용을 위해 Maxim 사의 3.3V Step-up 레귤레이터인 Max1678이 사용되었다. RF 통신을 위한 안테나는 PCB 안테나 또는 사용자의 선택사항으로 Dipole 안테나를 사용할 수 있다.

2. 센서 네트워크 소프트웨어

Nes-C는 C언어의 확장으로 TinyOS의 구현을 위해 제안되었다. TinyOS는 제한된 자원을 가진 센서 노드를 위해 제안된 event-driven 형태의 OS로 Atmega128L과 같이 128 kbyte의 롬과 4kbyte의 램에서도 수행될 수 있다. 일반적으로 센서 노드에 탑재되는 어플리케이션 프로그램은 TinyOS에서 기본적으로 제공하는 컴포넌트를 기반으로 Nes-C로 작성된다[11].

센서 네트워크에서 노드들간에 전송되는 TinyOS 메시지

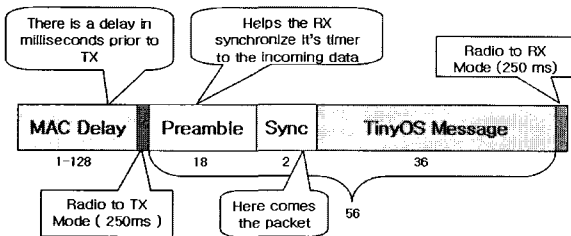


그림 2. TinyOS의 메시지 구조.
Fig. 2. Structure of TinyOS message.

```
typedef struct TOS_Msg
{
    uint16_t addr;
    uint8_t type;
    uint8_t group;
    uint8_t length;
    int8_t data[TOSH_DATA_LENGTH];
    uint16_t crc;
    uint16_t strength;
    uint8_t ack;
    uint16_t time;
} TOS_Msg;
```

그림 3. TOS_Msg 패키지의 구조.
Fig. 3. Structure of TOS_Msg packet.

패킷의 구성은 그림 2와 같으며 preamble은 데이터의 수신기와 클럭의 동기를 위한 것이고 프레임 싱크는 그림 3과 같은 TOS 메시지의 시작을 표시한다.

그림 3에서 addr 필드는 목적지 노드의 ID 또는 broadcast를 나타내며, group 필드는 네트워크에서 사용되는 센서 노드들의 채널을 표시한다. type 필드는 패킷이 받아졌을 때 TinyOS 레벨에서 불러지는 핸들러를 의미하며 length 필드는 data의 길이를 나타낸다. 일반적으로 TOS_Msg의 크기는 36byte를 넘지 않는다.

3. 센서 네트워크의 데이터 전송을 위한 Multi-hop 프로토콜

센서 네트워크에서는 RF의 거리 제약이 있기 때문에 광대한 영역에서의 데이터 계측 및 전송을 위해서는 Multi-hop 프로토콜이 요구된다. Multi-hop 프로토콜에는 기본적으로 flooding, gossiping, leach와 같은 기법들이 존재하며 이를 기반으로 다양한 프로토콜들이 연구되어지고 있다.

본 논문에서는 Multi-hop을 위한 응용 프로그램인 surge를 이용하여 데이터 계측에 사용하였다. Surge는 Tree 구조의 네트워크를 구성할 수 있는 응용 프로그램으로 전체 구조를 나타내면 그림 4와 같다.

TinyOS에서 surge 응용프로그램의 연결 구조를 살펴보면 다음과 같다. Surge의 Multi-hop 구조는 EWMAMultiHopRouter라는 컴포넌트로 구성이 가능하다. EWMAMultiHopRouter는 MultiHopEngineM컴포넌트와 연결되어 있고 MultiHopEngineM은 MultiHopEWMA컴포넌트와 연결되어 있다. 또한 MultiHopEWMA는 QueueSend, TimerC, GenericCommPromiscuous라는 컴포넌트로 이루어져 있다. 그림 5는 surge 알고리즘의 컴포넌트 구성도를 나타낸다.

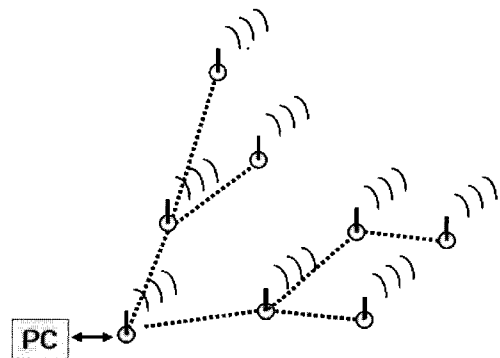


그림 4. Tree 네트워크의 구조.
Fig. 4. Structure of tree network.

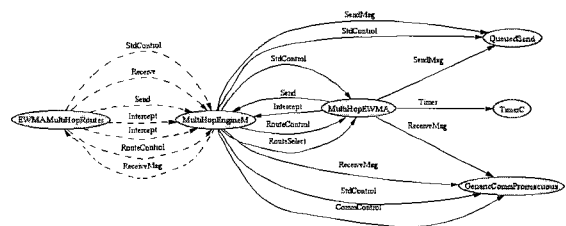


그림 5. Surge의 컴포넌트 구성도.
Fig. 5. Configuration of surge components.

III. PCA 기반 데이터 유효화 기법

PCA(Principle Component Analysis)는 다차원 변수들간에 존재하는 중요한 변량을 선형 결합을 통해 얻어지는 새로운 저차의 변수들로 표현하는 통계학적 모델링 기법으로 상관성이 큰 다변수 데이터의 처리를 가능케 한다. 이러한 PCA 기법은 다차원 데이터에 존재할지도 모르는 불확실성(데이터 오류 등)을 효율적으로 검출 및 보상할 수 있으며 따라서 본 연구에서는 센서 노드들로부터 싱크노드로 전송되어온 다차원 데이터에 존재하는 오류의 검출을 위해 사용하고자 한다.

n 개의 측정 변수들에 대한 m 개의 측정값을 $X(m \times n)$ 로 표현할 경우 이에 대한 주성분 분석 과정을 살펴보면 다음과 같다.

측정 행렬 X 에 대한 공분산 행렬은 다음과 같이 정의되며

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

공분산 행렬에 대한 고유 벡터 p_i 및 고유치 λ_i 는 다음과 같이 구해진다.

$$cov(X)P_i = \lambda_i P_i \tag{2}$$

만일 (2)에서 구한 고유값 중 가장 큰 고유값에 대한 고유벡터부터 r 개까지 내림차순으로 선택한 후, 다음의 식을 계산하면 r 개의 고유벡터만으로 복원된 x 의 추정치(\hat{x})를 구할 수 있다.

$$\hat{x} = \sum_{i=1}^r t_i \cdot p_i \quad r \leq n \tag{3}$$

여기서 $t_i = x \cdot p_i$ 이고 x 는 새로이 측정된 $n \times 1$ 차원의 측정 벡터이다.

(3)의 \hat{x} 는 r 개의 고유벡터만으로 복원된 x 의 추정치를 나타내며 r 개의 주성분만으로도 원래 x 의 정보를 효과적으로 표현할 수 있게 된다. 따라서 하나의 새로운 측정이 이루어질 경우, 새로운 측정 벡터 x 에 대한 모델 에러 e 는 다음과 같이 구할 수 있으며

$$e = (x - \hat{x}) \tag{4}$$

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

$$R = ee^T \tag{5}$$

따라서 (5)를 이용한다면 새로이 측정된 데이터 x 에 포함된 불확실성을 효과적으로 검출할 수 있게 된다.

IV. 센서네트워크 기반 데이터 유효화 기법

본 절에서는 앞서 고찰한 PCA 기법을 센서 네트워크상의 싱크 노드에서 수집된 다차원 측정 데이터에 적용함으로써 측정된 데이터의 유효화를 가능케 하고자 하며 제안된 시스템의 전체 구성은 다음과 같다.

그림 6과 같이 각 센서 노드들은 측정 데이터를 Multi-hop 라우팅 기법을 통해 원격의 임베디드 sink 노드로 전송

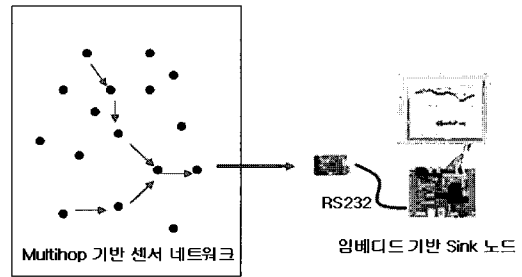


그림 6. 임베디드 시스템 기반 센서 네트워크의 구성.
Fig. 6. Configuration of sensor network based on embedded system.

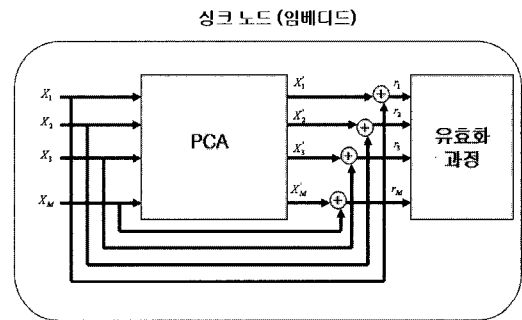


그림 7. 싱크노드에서의 데이터 유효화 기법.
Fig. 7. Data validation scheme in sink node.

하며 PCA가 구현된 임베디드 sink 노드는 앞서 고찰한 PCA 기반의 데이터 유효화 과정을 통해 측정된 데이터에 포함된 불확실성을 검출하게 된다. 싱크 노드에서 수행되는 데이터 유효화 과정을 나타내면 다음과 같다.

그림 7의 x 는 센서네트워크를 구성하는 센서 노드들로부터 전송되는 측정 데이터를 표현하며 PCA의 출력 결과인 x' 은 (3)에서의 복원과정을 거쳐 얻어진 결과를 의미한다. 또한 데이터 유효화 과정은 (4), (5)에 의한 임계치 검사 과정을 의미한다.

V. 제안된 시스템 성능 검토

본 연구에서는 제안된 PCA 기반의 데이터 유효화 기법의 유용성 확인을 위해 첫 번째로 임베디드 기반의 sink 노드를 갖는 Multi-hop 기반의 센서 네트워크 시스템을 구성하고, 두 번째로 sink 노드에 구현된 PCA의 정상 동작을 확인하기 위해 Matlab에서의 시뮬레이션 결과와 실제 sink 노드에서의 PCA수행 결과를 비교하여 보았다. 그림 8은 구성된 전체 시스템을 나타낸다.

1. Multi-hop 라우팅 기법을 이용한 센서 네트워크 구성

본 연구에서는 센서노드의 데이터 취득을 위한 Multi-hop 라우팅 기법으로 한백전자의 ZigbeX 어플리케이션에서 기본적으로 제공하는 surge 어플리케이션을 일부 수정하여 사용하였다. Surge의 올바른 동작을 확인하기 위해 20개의 ZigbeX 노드에 surge 어플리케이션을 포팅 한 후 surge-view라는 java 어플리케이션에서 동작을 확인해 보았다. Surge는 tree 토폴로지 형태를 갖는다.

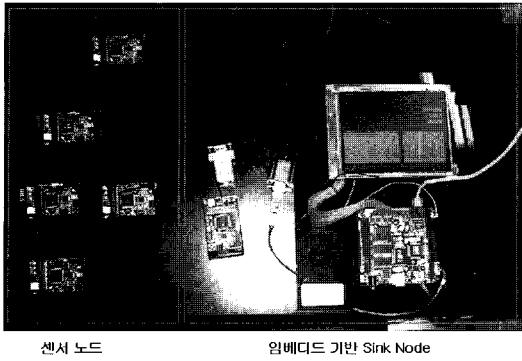


그림 8. 제안된 시스템의 구성.
Fig. 8. Configuration of the proposed system.

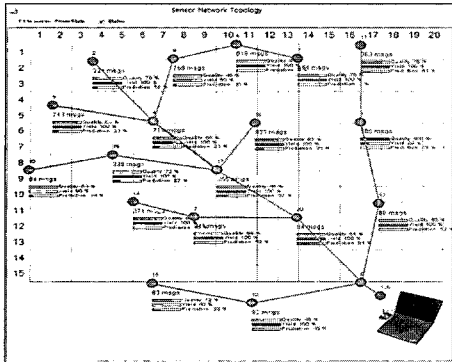


그림 9. Surge-view에서의 동작 화면.
Fig. 9. Operation screen under surge-view.

2. 임베디드 시스템에서의 PCA 구현

임베디드 시스템은 Intel사의 ARM칩인 Pxa255칩을 사용한 Hybus사의 X-hyper255A 보드를 사용하였다. X-hyper255A의 사양은 다음과 같다.

임베디드 sink node는 Multi-hop기반으로 구성된 센서 네트워크를 통해 전송된 데이터를 이용하여 PCA 과정을 거치게 된다. 이러한 임베디드 sink 노드에서의 PCA 구동은 다음과 같은 2가지 모드인 학습모드와 분석모드로 나뉘어진다.

2.1 학습모드

PCA를 이용한 유효화 과정을 위해서는 압축 및 복원을 위해 사용되는 저차원의 고유벡터행렬을 구하는 과정을 거쳐야 한다. 학습모드는 저차원의 고유벡터행렬을 구하는 과정을 말하며 세부 동작은 일정시간동안 구성된 센서네트워크를 통해 전송된 데이터를 저장해 두었다가 저장된 데이터를 이용하여 공분산 행렬을 계산하고 공분산 행렬의 고유값 및 고유벡터 행렬을 계산하게 된다. 계산된 고유값 중 가장 큰 몇 개의 고유값에 해당하는 고유벡터를 이용하여 압축 및 복원을 위해 사용하게 된다.

2.2 분석모드

분석모드는 데이터의 유효화 과정을 말하며 그 동작은 다음과 같다. 센서네트워크에서 새롭게 전송되는 계측 데이터와 학습모드에서 계산된 저차원의 고유벡터 행렬에 의한 복원데이터를 이용하여 얻은 결과값의 잔차를 계산하여

ITEM	Description
Process	Intel PXA255 400MHz
SDRAM	Samsung 32MByte
Flash	Intel strata flash 16MByte
Ethernet	CS8900A 10BaseT
Serial	2 Port
LCD	6.4" TFT LCD (640*480)
Touch	ADS7843(Touch screen)

그림 10. 임베디드 시스템의 사양.
Fig. 10. Specification of embedded system.

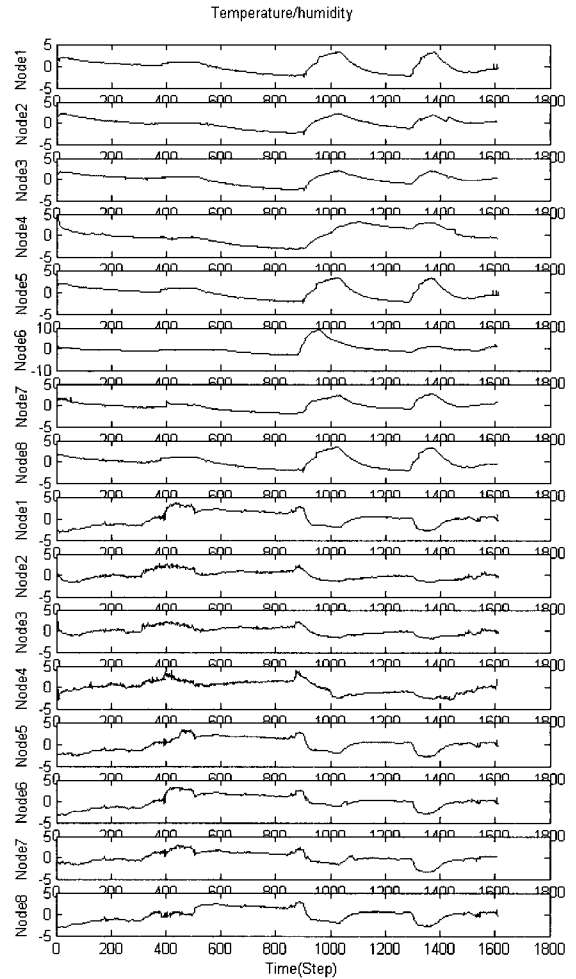


그림 11. 측정된 온/습도 데이터.
Fig. 11. Measured temperature/humidity data.

그 잔차가 정해진 임계치 이상이 되면 이상으로 인식하게 된다.

3. 임베디드 sink 노드의 정상 동작 및 유용성 확인

임베디드 sink 노드의 정상동작을 확인하기 위해 Matlab에서의 분석 결과와 실제 장치에서의 분석 결과를 비교해 보았다. 분석을 위해 사용된 데이터는 suege 어플리케이션이 포팅된 8개의 센서 노드를 통해 임베디드 장치로 전송되는 온/습도 데이터를 이용하였다. 취득된 데이터는 그림 11과 같다.

3.1 임베디드 sink 노드에서의 실시간 구현

그림 12는 실제 임베디드 장치에서 동작하는 프로그램을 나타낸다.

그림 12의 B 부분은 센서 노드에서 전송되는 온/습도가 포함된 데이터 패킷을 나타내며 C 부분은 데이터의 압축 및 복원을 위한 저차원의 추출 벡터를 나타낸다. 또한 A 부분은 8개의 센서 노드에서 측정된 데이터를 이용하여 PCA 기반의 유효화 과정을 통해 얻어진 잔차를 그래프로 나타낸 것이다.

임베디드 sink 노드에서의 실험은 2단계로 진행하였다. 1 단계로 학습모드에서는 2일간 측정된 온/습도 데이터로부터 얻어진 16개의 고유값 중 가장 큰 8개의 고유값을 이용하여 8 차원의 고유벡터를 추출하였고, 2단계로 분석모드에서 새롭게 입력되는 계측데이터와 PCA 기반의 유효화 과정을 통한 결과 값을 이용하여 데이터 속에 포함된 불확실성을 검출하도록 하였다.

전송되어 온 데이터에 포함된 불확실성의 검출가능성을 확인하기 위해 5번째 센서 노드를 정상적인 데이터를 얻을 수 있는 위치에 설치해 두었다가 인위적으로 해당센서의 값을 변경하였다.

그림 13은 D 시점에서 해당 센서의 측정값에 불확실성이 포함되었을 경우, PCA 기반의 데이터 유효화 알고리즘의 응답 특성을 나타낸 것으로 효율적인 무효 데이터의 검출이 가능함을 확인할 수 있다.

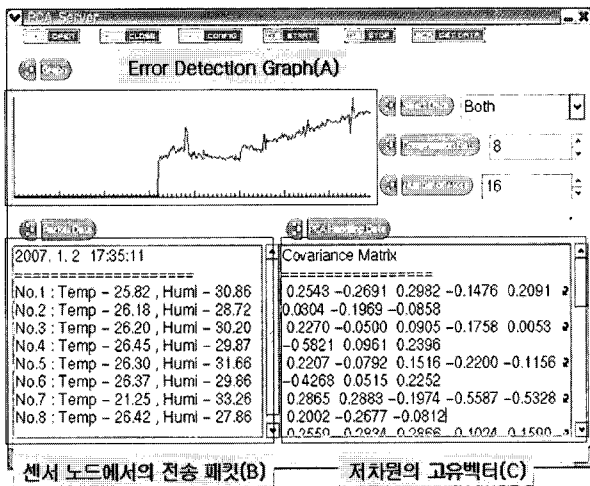


그림 12. 임베디드 sink 노드에서의 동작 화면.
Fig. 12. Operation screen on embedded sink node.

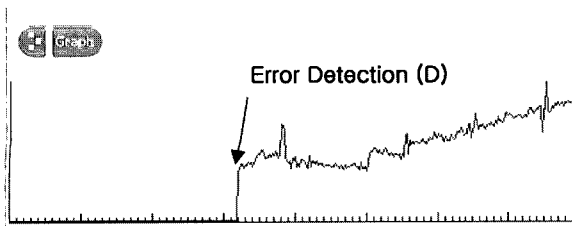


그림 13. Error detection 화면.
Fig. 13. Error detection screen.

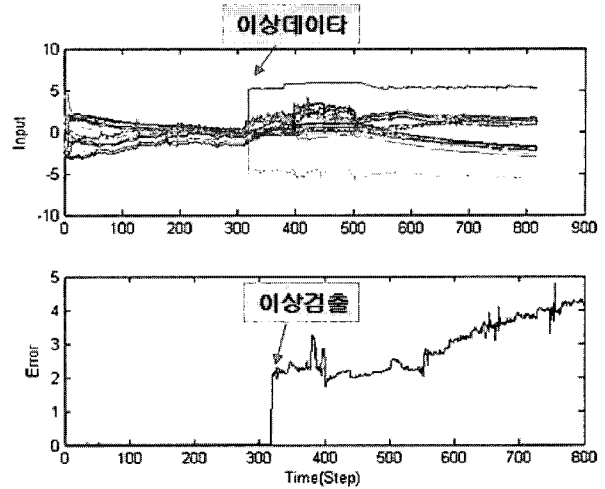


그림 14. 시뮬레이션 결과.
Fig. 14. Simulation result.

3.2 Matlab을 이용한 시뮬레이션

상기 임베디드 시스템에서 구현된 PCA 기반의 데이터 유효화 기법의 유용성 확인을 위해 그림 11과 같은 온/습도 데이터에 대한 시뮬레이션을 수행하였다. 시뮬레이션은 Matlab에 의해 수행되었으며 임베디드 sink 노드에서 얻어진 온/습도 데이터를 이용하여 공분산 행렬을 구했으며 16개의 고유값 중 가장 큰 8개를 이용하여 8차원의 고유벡터를 얻었다. 두 번째로 임베디드 sink 노드의 분석모드에 사용된 데이터를 일정간격 저장한 후 이것을 이용하여 데이터에 포함된 불확실성 검출의 시뮬레이션을 수행하였다. 그림 14는 Matlab상에서 데이터에 포함된 불확실성의 검출을 나타낸 것으로 그림 13의 실시간 응답특성과 정확히 일치함을 확인할 수 있었다.

VI. 결론

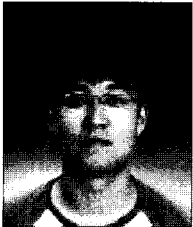
본 연구에서는 다수의 센서 노드들로부터 원격의 싱크 노드로 전송된 데이터에 포함될 수 있는 각종 불확실성을 효과적으로 검출할 수 있는 PCA기반의 데이터 유효화 기법을 제안하고 실제 임베디드 기반 싱크노드상에서 실시간 구현 및 Matlab에 의한 시뮬레이션을 수행하였다. 수행결과 PCA 기반의 데이터 유효화 기법은 전송된 데이터에 포함되는 불확실 요인을 효율적으로 검출할 수 있음을 확인하였다.

참고문헌

[1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP2001)*, Salt Lake City, Utah, May 2001.

[2] S. Chessa and P. Santi, "Crash faults identification in wireless sensor networks," *Computer Communications*, vol. 25, no. 14, pp. 1273-1282, Sept. 2002.

- [3] K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiley & Sons, Inc., 1996.
- [4] R. Dunia, and S. J. Qin, "Identification of faulty sensors using principle component analysis," *AIChE J.*, vol. 42, no. 10, pp. 2797-2812, 1996.
- [5] H. Vedam and V. Venkatasubramanian, "Automated interpretation of PCA-based process monitoring and fault diagnosis using digraphs."
- [6] R. Dunia, "Identification of faulty sensors using principle component analysis," *AIChE J.*, 42(10), 2797-2812, 1996.
- [7] W. B. Heinzelman, *Application-specific Protocol Architectures for Wireless Networks*, Ph.D. dissertation, Mass. Inst. Tech., Cambridge, 2000.
- [8] P. Levis, N. Lee, M. Welsh, and D. Culler, "TOSSIM: accurate and scalable simulation of entire TinyOS applications," *UC Berkeley*. June 26. 2003.
- [9] P. Levis and D. Gay, Tinyos design patterns, <http://www.cs.berkeley.edu/~pal/tinyos-patterns>, 2004.
- [10] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *In Proceedings of the Hawaii International Conference on System Sciences*, January 2000.
- [11] P. Levis, "The TinyScript language," July, 2004.
- [12] P. Levis, "Ad-Hoc routing component architecture," February 5, 2003.
- [13] Geoff Martin, "An evaluation of Ad-hoc routing protocols for wireless sensor networks," May, 2004.
- [14] B. Kumar, K. and Jaffe, J. M. "Routing to Multiple destinations in computer networks," *IEEE trans. Commun. COM-31*, Mar. 1983.



육 의 수

2004년 군산대학교 공과대학 학사.
2006년 군산대학교 대학원 석사. 2006
년~현재 군산대학교 대학원 박사과정.
관심분야는 고장진단, 공장 자동화,
Sensor Network.



김 성 호

1984년 고려대학교 공과대학 학사.
1986년 고려대학교 대학원 석사. 1991
년 고려대학교 대학원 박사. 1995
년~1996년 Japan Hiroshima University,
Post-DOC. 1997년~현재 군산대학교 전
자정보공학부 교수. 관심분야는 고장
진단, 공장 자동화, Sensor Network.