

<http://dx.doi.org/10.7236/JIIBC.2013.13.3.33>

JIIBC 2013-3-5

스마트폰 센싱을 위한 손실 데이터 추정 모델

An Estimation Model of Missing Data for Smart Phone Sensing

민 홍*, 허준영**

Hong Min, Junyoung Heo

요 약 스마트폰에 탑재된 다양한 종류의 센서들을 활용하여 사용자의 상태나 사회활동 및 주변 환경을 모니터링 하는 스마트폰 센싱 시스템에서 특정 지역의 데이터가 손실되는 문제는 피할 수 없다. 다수의 사용자를 대상으로 사전에 정의해 놓은 조건이 만족할 때 센서로부터 측정된 값을 서버로 전송하는 기회기반 센싱 기법에서는 이러한 데이터 손실 문제가 더 심화된다. 본 논문에서는 수집된 데이터의 품질 저하 문제를 해결하기 위해 스마트폰 센싱의 특성을 고려한 손실 데이터 추정 모델을 제안한다. 제안된 추정 모델에서는 데이터의 시공간적 상관관계를 고려할 뿐만 아니라 신뢰도가 높은 데이터를 제공하는 참여자의 우선순위를 높임으로써 향상된 추정 값을 도출하도록 설계하였다. 또한 실험결과를 통해 본 논문에서 제안한 기법이 기존의 기법들에 비해 높은 신뢰도를 보이는 것을 알 수 있었다.

Abstract Smart phones that are equipped with various types of sensors can monitor human beings, and their social activities and environments. Smart phone sensing systems are inevitable to lose sensing data at a certain region. It is more severe effect on opportunistic sensing because this sensing scheme is designed to read values from sensors when the state of numerous users meets pre-defined conditions. In this paper, we suggested an estimation model of missing data considering features of smart phone sensing to solve lower quality of collected data. This proposed model does not only reflect a temporal and spatial correlation, but also give high priority to participants who provide high quality data to improve the accuracy of estimated values. The experimental results show that our scheme is more accurate than previous scheme.

Key Words : Smart phone sensing, Data missing region, Estimation model, Opportunistic sensing, Participatory sensing

1. 서 론

최근 스마트폰은 휴대용 통신기기로서의 역할뿐만 아니라 탑재되어 있는 다양한 종류의 센서들을 바탕으로 사용자의 상태와 주변 환경을 모니터링 할 수 있는 장치로서의 역할도 수행할 수 있다^[1]. 다수의 값싼 기기에 의존했던 기존의 무선 센서 네트워크와는 달리 성능

이 뛰어나고 사용자에 의해 원하는 목적지로 이동이 가능한 스마트폰을 활용하여 데이터를 수집하는 스마트폰 센싱 분야는 앞으로 사용자의 사회 활동과 연관되어 많은 응용 분야에서 사용될 것으로 전망된다^[2, 11].

스마트폰 센싱 분야는 발전 가능성에 비해 현재 진행되고 있는 연구들은 아직 초기 단계를 벗어나지 못하고 있으며^[3] 정적인 무선 센서 네트워크와는 달리 동적인

*정희원, 호서대학교, 모바일시스템공학

**정희원, 한성대학교, 컴퓨터공학과

접수일자 : 2013년 4월 8일, 수정완료 : 2013년 5월 11일

게재확정일자 : 2013년 6월 14일

Received: 8 April 2013 / Revised: 11 May 2013 /

Accepted: 14 June 2013

*Corresponding Author: jyheo@hansung.ac.kr

Dept. of Computer Engineering, Hansung University

요소를 포함하고 있기 때문에 기존의 연구결과를 그대로 적용하기 힘들다는 특성을 가지고 있다^[4]. 노드의 이동성과 더불어 한 영역을 서로 다른 플랫폼을 갖는 장치들이 서로 다른 시간대에 모니터링하기 때문에 센싱된 데이터의 시공간적 상관관계가 많이 떨어진다. 이는 모든 데이터를 서버 단에서 수집한 후에 손실된 데이터를 추정하는 기존 모델들의 정확성을 악화시켜 수집된 데이터의 품질 저하 문제를 유발한다^[5].

본 논문에서는 이러한 센싱 데이터의 품질 저하 문제를 해결하기 위해 스마트폰 센싱의 특성을 고려한 손실 데이터 추정 모델을 제안한다. 제안된 모델에서는 클러스터링 기법을 활용하여 센싱 대상이 되는 타겟 필드를 작은 영역(클러스터)들로 분할하고 각 영역에서 클러스터 헤드를 선출하여 다른 멤버 노드와 역할을 분리한다. 클러스터 헤드는 일정 기간 동안 자신의 영역 내에 머물면서 시간의 변화에 따라 센싱 데이터의 값의 변화를 추적 한다. 또한 클러스터 내에 특정 영역에서 데이터의 손실이 발생할 경우 인접 지역의 멤버 노드들의 센싱 값을 바탕으로 손실된 데이터 값을 추정한다. 이러한 계층적 구조를 통해 서버단에서의 프로세싱 부하를 줄일 수 있으며 데이터 손실 영역에 대한 시공간적 상관관계를 향상시킴으로써 전반적인 데이터의 품질을 개선시킬 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 스마트폰 센싱 기법과 데이터 손실 영역에 대한 설명을 통해 해결해야할 문제를 정의하며, 3장에서는 제안한 손실 데이터 추정 모델에 대해 서술한다. 4장에서는 제안한 방법의 실험 및 결과를 살펴보고, 5장에서는 결론을 맺는다.

II. 데이터 손실 영역

1. 스마트폰 센싱 기법

스마트폰 센싱에서는 센싱 데이터 수집 과정에서 장치 사용자의 역할에 따라 두 가지의 센싱 기법으로 구분하고 있다^[6]. 먼저 참여기반 센싱 기법은 시스템 관리자가 장치 사용자들에게 보상을 통해 참여를 독려하여 응용에서 요구하는 목적을 만족시키는 접근 방법이다. 사용자의 역할이 데이터 수집에 있어 중요한 영향을 끼치기 때문에 기밀성을 유지하여 사용자의 정보 공개에 대한 거부감을 줄이는 노력들이 부가적으로 수행되어야

한다.

다음으로 기회기반 센싱 기법은 사전에 정의된 조건들이 부합하는 상황에 우연히 기기가 노출될 경우 자동으로 센싱된 데이터를 서버로 전송하는 접근 방법이다. 이러한 기회기반 접근 방법은 사용자와 정보 공개에 대한 사전 합의가 없기 때문에 수집되는 정보의 종류가 제한적이고 센서가 정확한 데이터를 수집할 수 있을 만큼의 노출 시간을 보장할 수 없기 때문에 수집된 데이터의 품질이 저하되는 문제가 있다. 표 1은 두 센싱 기법에 대한 차이점을 비교하고 있다.

표 1. 참여기반 센싱과 기회기반 센싱 비교
Table 1. Comparison between participatory sensing and opportunistic sensing

	참여기반 센싱 (Participatory sensing)	기회기반 센싱 (Opportunistic sensing)
네트워크 규모	소규모	대규모
데이터 품질	높음	낮음
보상 정책	필요	불필요
사용자 편의성	낮음	높음
시스템 설계의 복잡성	낮음	높음

2. 데이터 손실 영역 문제

스마트폰을 활용하여 특정 집단이나 지역을 모니터링 하는 대규모 센싱 시스템은 시간과 비용의 효율성 때문에 기회기반 센싱에 의존할 수밖에 없다. 이러한 기회기반 센싱의 경우 시간의 흐름이나 지역적 특성에 따라 일부 영역의 데이터가 손실되는 문제가 발생한다. 그림 1은 시공간적으로 발생할 수 있는 데이터 손실 문제를 보여준다.

그림 1의 (a)와 (b)은 시간의 흐름에 따라 데이터 손실 영역이 발생하는 상황을 설명하고 있다. 사용자들의 이동으로 인해 특정 샘플링 시점에서 센싱 참여 노드가 위치하지 않는 영역이 발생하고 이는 시간과 공간에 따라 변화가 커진다. A 영역의 경우 두 번의 샘플링 시점에서 연이어 데이터 값이 손실되었기 때문에 시간적인 상관관계를 통한 데이터 값 추정이 힘들다. B 영역의 경우 상당한 범위에서 데이터 값이 손실 되어 공간적인 상관관계가 약해지고 이로 인해 주변의 값으로부터 데이터를 추정하는 것이 어렵다.

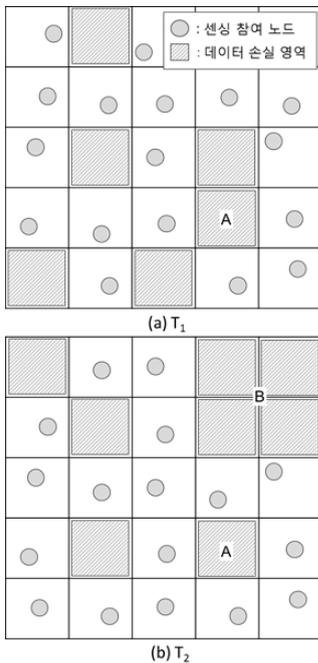


그림 1. 데이터 손실 영역 발생의 예
Fig. 1. An example of data missing region

III. 손실 데이터 추정 모델

1. 가정

데이터 손실 영역 발생 시 이를 복구하는 과정과 손실 데이터 값을 추정하는 모델은 다음과 같은 가정 하에서 설계되었다.

- 두 센싱 기법간의 데이터 수집 성공 확률은 Lane 등^[7]이 제안한 모델에 기반을 둬
- 공간적 상관관계를 반영하기 위한 추정 모델은 Pan 등^[8]이 제안한 모델에 기반을 둬
- 클러스터링 구성 및 관리 알고리즘은 고려하지 않음
- 네트워크 통신상의 오류는 고려하지 않음
- 모든 노드는 일정한 속도로 이동

두 센싱 기법에서 데이터 손실 확률은 스마트폰에 탑재된 센서가 특정 시간과 장소에서 일정기간 노출되는 것을 보장해줄 수 있는지 여부에 의해 결정된다. Lane 등은 이러한 데이터 수집 성공 확률을 간단한 모델을 통해 제안하고 있다.

Pan 등은 무선 센서 네트워크에서 클러스터링 기법을 사용하여 멤버 노드에서 데이터의 손실이 발생했을 때 이를 클러스터 단위로 손실된 데이터를 복구하는 추정 모델을 제안하였다. 본 논문에서 제안된 추정 모델은 기존의 모델을 기반으로 하여 스마트폰 센싱에 적합하도록 수정되었다.

클러스터를 구성하기 위해서는 클러스터 헤드를 선출하는 것이 선행되어야 한다. 제안된 모델에서 클러스터 헤드는 참여기반 센싱 기법을 사용하게 되는데, 응용에 적합한 참여자를 모집하는 방법은 Reddy 등^[9]이 제안하는 방법을 활용할 수 있다. 또한 클러스터를 구성하고 관리하는 알고리즘은 다양한 무선 통신 수단과 위치기반 서비스를 지원하는 스마트폰에서 간편하게 설계 가능하고 본 논문의 범위를 벗어나기 때문에 고려하지 않는다.

2. 시스템 개요

본 논문에서는 특정 집단이나 환경을 모니터링 하는 기존의 대규모 센싱 시스템이 기회기반 센싱에 의존했던 것과는 달리 참여기반 접근 방법과 기회기반 접근 방법을 모두 활용하여 데이터의 품질을 향상시킬 수 있는 클러스터링 기법을 제안한다. 그림 2는 제안된 클러스터링 기법을 보여준다.

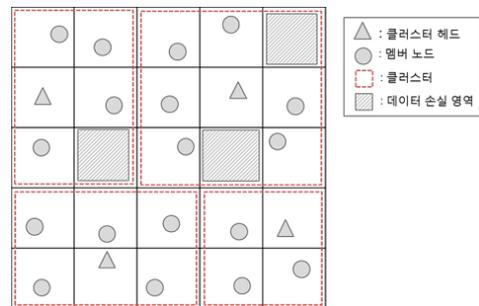


그림 2. 참여 센싱 기반 클러스터링 기법
Fig. 2. A clustering scheme based on participatory sensing

클러스터 내의 멤버 노드들은 기회기반 센싱 기법으로 설계되었기 때문에 클러스터 헤드가 데이터를 요청하는 메시지를 받을 경우 수락 여부에 따라 센싱 값을 전송하거나 무시할 수 있다. 클러스터 헤드의 요청을 수락한 경우에도 멤버 노드의 상태가 사전에 정의된 조건에 부합하지 않는다면 센싱 데이터를 전송할 수 없다.

반면에 클러스터 헤드는 참여기반 센싱 기법으로 설계되었기 때문에 매 샘플링 시점에 자신에게 할당된 영역의 센싱 값을 측정한다. 또한 멤버 노드들에게 센싱 데이터의 타입과 필요한 조건들을 포함한 요청 메시지를 전송하고 요청 메시지를 수락한 멤버 노드들로부터 센싱 데이터를 전송받아 이를 병합하여 서버에 전송하는 역할을 수행한다. 이 때, 멤버 노드로부터 데이터를 받지 못하여 자신의 클러스터 내에 데이터 손실 영역이 발생하면 3절에서 설명한 손실 데이터 추정 모델을 사용하여 손실된 데이터를 보정한 후에 서버로 수집된 데이터를 전송한다. 이러한 기법을 통해 손실 데이터가 발생한 시점에서 바로 보정이 가능하여 데이터의 신뢰도를 높일 수 있고, 추정 모델에 사용될 데이터의 양이 줄어들어 서버의 프로세싱 부하도 줄일 수 있다.

3. 손실 데이터 추정 모델

본 논문에서 제안한 손실 데이터 추정 모델은 센싱된 데이터의 시공간적 상관관계를 활용하여 데이터 손실 영역의 센싱 값을 계산한다. 수식 (1)은 매 라운드(r) 마다 클러스터 내에서의 데이터 손실 영역 발생 시 헤드 노드에서 수행하는 추정 방법을 보여준다.

$$\widehat{E}_r^i = \omega \times \widehat{T}_r^i + (1 - \omega) \times \widehat{S}_r^i \quad (1)$$

주어진 식에서 \widehat{T}_r^i 은 시간적 상관관계를 고려한 예측 값을 나타내고 \widehat{S}_r^i 은 공간적 상관관계를 고려한 예측 값을 나타낸다. 두 예측 값 사이의 가중치(ω)는 수식 (2)을 통해 계산할 수 있다. 가중치 함수에서 $d(N_i, M)$ 은 데이터 손실 영역(M)과 멤버 노드(N_i) 사이의 거리를 의미한다. 가중치 값은 데이터 손실 영역 주변에 노드가 많을 경우 작은 값을 가지게 되고 이는 추정 모델에서 공간적 상관관계를 높게 반영하는 것을 의미한다.

$$\omega = \frac{\sum_{i=1}^{n-1} D_i}{n-1}, \quad (2)$$

$$D_i = \begin{cases} 0, & d(N_i, M) \leq d_{\max} \\ 1, & d(N_i, M) > d_{\max} \end{cases}$$

시간적 상관관계는 수식 (3)에서 보이는 것처럼 데이터 손실 영역을 포함하고 있는 클러스터의 헤드 노드(H_i)의 연속적인 측정값을 활용하여 추정 모델에 반영하였다. 즉 단위 시간 동안의 변화량에 샘플링 주기(I)를 곱하는 방법을 사용한다.

$$\widehat{T}_r^i = \frac{V_r^{H_i} - V_{r-1}^{H_i}}{t_r - t_{r-1}} \times I \quad (3)$$

다음으로 공간적 상관관계를 고려하기 위해 수식 (4)에서 보이는 것처럼 선형 회귀 모델 (Linear Regression Model)을 기반으로 도출된 추정치를 사용한다. 제안한 추정 모델의 신뢰도를 결정하는 중요한 요인은 페어링 ($V_r^{x_i}, V_r^{y_i}$) 방법에 있다. 페어링 과정에서 클러스터 헤드와 데이터 손실 영역에 인접한 노드들의 선택 빈도를 높임으로써 오차발생을 최소화하고자 하였다.

$$\widehat{S}_r^i = \widehat{\alpha} + \widehat{\beta} \times V_r^{H_i}, \quad \widehat{\alpha} = \overline{V_r^y} - \widehat{\beta} \times \overline{V_r^x},$$

$$\widehat{\beta} = \frac{\sum_{i=0}^n (V_r^{x_i} - \overline{V_r^x})(V_r^{y_i} - \overline{V_r^y})}{\sum_{i=0}^n (V_r^{x_i} - \overline{V_r^x})^2}, \quad (V_r^{x_i}, V_r^{y_i}) \quad (4)$$

IV. 실험 및 결과

본 절에서는 기존의 기법들과 제안된 추정 모델의 오류 발생 정도를 비교하여 신뢰도를 검증한다. 이를 위한 평가 기준으로 평균 제곱근 오차(Root Mean Square Error; RMSE)를 사용한다.

$$RMSE = \sqrt{\frac{\sum (\widehat{E}_r^i - E_r^i)^2}{\text{Num of Data Missing Region}}} \quad (5)$$

실험을 위해서 실측 데이터와 함께 다양한 환경 변화 모델을 적용할 수 있는 SIDnet-SWANS^[10] 시뮬레이터를 사용하였다. 표 2는 실험에 사용한 파라미터 목록과 그 값을 보여준다.

표 2. 실험 파라미터

Table 2. Simulation Parameters

Parameter	설정 값
필드 크기	500 X 500
노드의 개수	200
참여 센싱 기반 노드 개수	40 (20%)
샘플링 주기 (T)	5 min
데이터 손실 영역 비율	20%
참여 센싱 기법의 성공률	2%
기회 기반 기법의 성공률	0.5%

그림 3은 매 라운드 마다 기회기반 센싱 기법(*opp*)과 20%의 참여센싱 기반 노드를 할당한 제안 기법(*par-20%*)의 평균제곱근오차를 비교한 결과이다. 기존의 기회기반 센싱에 의존하는 경우 매 라운드 마다 오차 값의 변화가 급격한 것을 알 수 있다. 반면에 기회기반 센싱과 참여기반 센싱을 활용한 제안기법의 경우에는 클러스터 헤드의 보정효과 때문에 추정 값과 실측 값 사이의 오차가 작고 라운드 수가 증가하더라도 안정적인 상태를 유지하는 것을 알 수 있다.

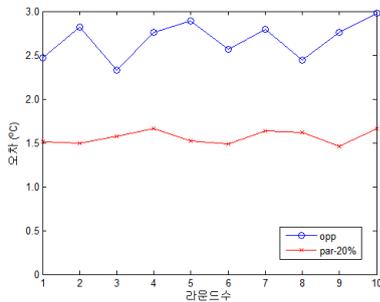


그림 3. 기회기반 센싱과 제안 기법의 평균제곱근오차 비교
Fig. 3. RMSE comparison between opportunistic sensing and our proposed scheme

그림 4와 5는 시뮬레이터를 사용하여 데이터를 수집하고자하는 센싱 영역이 균등하게 분포되어 있을 때와 임의로 분포되어 있을 때의 오차를 비교한 것이다.

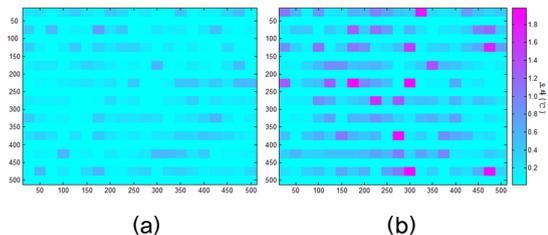


그림 4. 센싱 영역 분포가 균등할 때 오차 비교
Fig. 4. Error comparison in uniform deployment of sensing regions

그림 4의 (a)은 제안기법의 결과를 (b)은 기존의 기회기반 센싱을 했을 때의 결과를 보여준다. 제안 기법의 경우 클러스터링을 통한 공간적 상관관계와 참여 센싱 기반 헤드 노드를 통한 시간적 상관관계가 추정 모델에 고려되어 있기 때문에 전반적으로 오차 범위가 작은 것을 알 수 있다. 그러나 기존의 기회기반 센싱을 이용할 경우에는 오차가 큰 핫스팟이 많이 발생하는 것을 알 수 있다.

센싱 영역 분포가 랜덤할 때는 균등한 분포에 비해 오차가 증가하는데 제안 기법 보다는 기존의 기회기반 센싱 기법을 사용할 때 더 큰 영향을 받는 것을 알 수 있다.

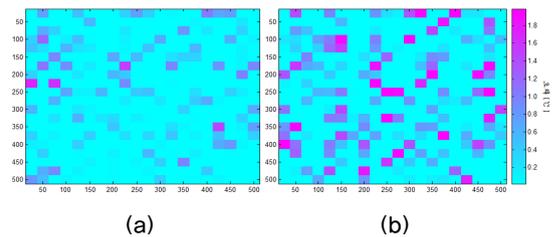


그림 5. 센싱 영역 분포가 랜덤할 때 오차 비교
Fig. 5. Error comparison in random deployment of sensing regions

V. 결론

다양한 종류의 센서들을 탑재하고 있는 스마트폰은 모바일 센싱 시스템 분야에 큰 변화를 가져오고 있다. 대중화된 스마트폰을 통해 수집된 정보들은 각종 응용 분야에 전파되어 인간 생활을 향상 시키는데 많은 도움을 줄 것으로 예상된다. 이러한 응용들을 위해 기반 시설을 설계하고 시스템을 운영하는 방법들은 최근 많은 연구가 진행되고 있다. 그러나 높은 비용과 시간을 소모하여 수집된 데이터의 품질을 높이려는 연구는 아직 부족하다.

본 논문에서는 기회기반 센싱 기법으로 스마트폰 센싱을 운용했을 때 발생할 수 있는 데이터 손실 영역 문제를 정의하고 이를 보정하기 위한 추정 모델을 제안했다. 제안된 추정 모델은 스마트폰 센싱의 특성을 고려했을 뿐만 아니라 수집된 데이터의 시공간적 상관관계를 활용하여 추정 값의 신뢰성을 높이고 클러스터링 기법을 통해 문제가 발생한 지점에서 복구가 이루어지기 때문에 서버의 부하를 줄일 수 있다는 장점이 있다. 성능 평가 결과를 통해서 제안된 기법이 기존의 기법에 비해 평균제곱근오차를 줄일 수 있으며 보다 안정적인 데이터 품질을 유지할 수 있음을 검증하였다.

References

- [1] A. T. Campbell, "From Smart to Cognitive Phones," IEEE Pervasive Computing, Vol. 11, No.3, pp.7~11, Jun. 2012.
- [2] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, Q. Arshad, "Mobile Phone Sensing Systems: A Survey," IEEE Communications Surveys & Tutorials, Vol.15, No.1, pp.402~427, Apr. 2013.
- [3] Y. Kang, and M. Hong, "Sensory Data Processing by Using Hadoop Framework," Journal of Korean Institute of Information Technology, Vol. 11, No. 2, pp. 169~174, Feb. 2012.
- [4] R. Gant, F. Ye, H. Lei, "Mobile Crowdsensing: Current State and Future Challenges," IEEE Communications Magazine, Vol. 49, No. 11, pp.32~39, Nov. 2011.
- [5] J. Kim, "A Layer-based Dynamic Unequal Clustering Method in Large Scale Wireless Sensor Networks," Journal of the Korea Academia-Industrial cooperation Society, Vol. 13, No. 12, pp.6081~6088, Dec. 2012.
- [6] T. Das, P. Mohan, V. N. Padmanabhan, R. Ramjee, A. Sharma, "PRISM: platform for remote sensing using smartphones," The 8th international conference on Mobile systems, applications, and services, pp.63~76, Jun. 2010.
- [7] N. Lane, S. Eisenman, M. Musolesi, E. Miluzzo, A. Campbell, "Urban Sensing Systems: Opportunistic or Participatory?" The 9th Workshop on Mobile Computing Systems and Applications, pp. 11~16, Feb. 2008.
- [8] L. Pan, J. Li, "K-Nearest Neighbor Based Missing Data Estimation Algorithm in Wireless Sensor Networks," Wireless Sensor Network, Vol. 2, No. 2, pp. 115~122, Feb. 2010.
- [9] S. Reddy, D. Estrin, M. Srivastava, "Recruitment Framework for Participatory Sensing Data Collections," The 8th International Conference on Pervasive Computing, pp. 138~155, May 2010.
- [10] O. Ghica, G. Trajcevski, P. Scheuermann, Z. Bischof, N. Valtchanov, "SIDnet-SWANS: a simulator and integrated development platform for sensor networks applications, The 6th ACM conference on Embedded networked sensor systems," pp. 385~386, Nov. 2008.
- [11] C. Kim, N. Kim, "Collision Avoidance Power Control of Carrier Sensing Zone for Energy Efficiency in Wireless Sensor Network," The Journal of The Institute of Internet, Broadcasting and Communication, Vol. 11, No. 4, pp. 53~60, Aug. 2011

※ 본 연구는 한성대학교 연구장려금 지원과제임.

저자 소개

민 흥(정회원)



- 2004년 : 한동대학교 전산과학 졸업(학사).
- 2011년 : 서울대학교 컴퓨터공학부 졸업(박사).
- 2013년 ~ 현재 : 호서대학교 모바일 시스템공학과 조교수.

<주관심분야 : 운영체제, 무선 센서 네트워크, 스마트폰 센싱, 임베디드 시스템, 결합허용 시스템>

허 준 영(정회원)



- 1998년 : 서울대학교 컴퓨터공학과 졸업(학사).
- 2009년 : 서울대학교 컴퓨터공학부 졸업(박사).
- 2009년 ~ 현재 : 한성대학교 컴퓨터 공학과 조교수.

<주관심분야 : 운영체제, 무선 센서 네트워크, 임베디드 시스템, 결합허용 시스템>