

## 효율적인 실내 측위를 위한 KNN/SVM 알고리즘

강일우\*, 로네쉬\*, 전성민\*, 박선\*\*, 이성호\*\*, 나영화\*\*, 배진수\*\*\*, 정민아\*\*\*\*,  
이연우\*\*\*\*, 이성로\*

\*목포대학교 정보전자공학과

\*\*목포대학교 정보산업연구소

\*\*\*세종대학교 정보통신공학과

\*\*\*\*목포대학교 컴퓨터공학과

\*\*\*\*\*목포대학교 정보통신공학과

e-mail:kiw2673@naver.com

### Optimized KNN/SVM Algorithm for Efficient Indoor Location

Il-Woo Kang\*, Ronesh Sharma\*, Sun Park\*\*, Young-Hwa Na\*\*,  
Jinsoo Bae\*\*\*, Min-A Jung\*\*\*\*, Yeonwoo Lee\*\*\*\*\*, Seong-Ro Lee\*

\*Dept of Information & Electronics Engineering, Mokpo National University

\*\*Research Institute of Information Science & Engineering, Mokpo National University

\*\*\*Dept of Information & Communication Engineering, Sejong University

\*\*\*\*Dept of Computer Engineering, Mokpo National University

\*\*\*\*\*Dept of Information & Communication Engineering, Mokpo National University

#### 요 약

현재 측위에 대한 측정 대상이 점점 작아지면서, 그에 따른 정확도 까지 높아지고 있다. 실내 측위에 관한 기술은 대표적으로 단말기의 수신신호의 세기방식인 RSS(Received Signal Strength), 수신신호의 도달시간 방식 TOA(Time of Arrival), 수신 신호의 도달 시간차 방식 TDOA(Time Difference of Arrival), 수신신호의 입사각 방식인 AOA(Angle of Arrival) 등 여러 가지 기술이 활발히 진행되고 있다. 본 논문은 특수 장비를 사용하지 않고, 무선 네트워크 기반의 실내 측위 중에 정확도가 높은 Fingerprinting 방법을 택하였다. WLAN 기반 실내측위에 가장 많이 사용되는 KNN은 k개의 이웃수와 RP의 수에 따라 민감하다. 본 논문에서는 KNN 성능을 향상 시키기 위해 SVM 이용하여 SNR 데이터를 군집화를 적용한 KNN과 SVM을 혼합한 알고리즘을 제안하였다. 제안한 알고리즘은 신호잡음비 데이터를 KNN 방법에 적용하여 k개의 RP를 선택한 후 선택된 RP의 신호잡음비를 SVM에 적용하여 k개의 RP를 군집하여 분류한다. 실험 결과 위치 오차가 2m이내에 KNN/SVM 혼합 알고리즘이 KNN 알고리즘보다 성능이 우수하다.

#### 1. 서론

현재 누구나 사용하고 있는 측위의 대표적인 예가 GPS(Global Positioning System) 라고 할 수 있다. 하지만 이 기술은 위성을 통하여 실외에 적용되는 기술이고, 실내에는 사용하지 못하는 단점을 가지고 있다. 측위에 대한 측정 대상이 점점 작아지면서, 그에 따른 정확도 까지 높아지고 있다. 실내 측위에 관한 기술은 대표적으로 단말기의 수신신호의 세기방식인 RSS(Received Signal Strength), 수신신호의 도달시간 방식 TOA(Time of Arrival), 수신 신호의 도달 시간차 방식 TDOA(Time Difference of Arrival), 수신신호의 입사각 방식인 AOA(Angle of Arrival) 등 여러 가지 기술이 활발히 진행되고 있다. 측위에 대한 관심이 높아짐에 따라 여러 가지 방식이 발전 되고 있는데, 최근 들어 무선 랜 노트북이나 넷 북 그리고 스마트 폰 사용이 늘어남에 따라 건물

내 무선 랜 사용도가 늘어나고 있다. 본 논문은 특수 장비를 사용하지 않고, 무선 네트워크 기반의 실내 측위 중에 정확도가 높은 Fingerprinting 방법을 택하였다. Fingerprinting 방식에 속하는 기술은 다양하지만 가장 기본적으로 널리 사용되는 기술은 KNN이다. KNN알고리즘은 무선 랜 기반 실내 측위에 간단하게 적용가능하고 성능적인 측면에 가장 많이 사용되고 있으나, k개의 이웃수와 RP의 수에 따라 민감하고 광범위하고 복잡한 실내 환경에서 거리오차가 많이 발생하여 실제 환경에서 최적의 위치를 찾기가 어렵다. 이러한 문제를 보완하기 위해 Fingerprinting 방식에서 가장 많이 사용되는 KNN방법으로 분류되어진 k개의 그룹 데이터를 SVM 알고리즘을 사용하여 최적의 위치를 식별해내는 KNN/SVM 혼합 알고리즘을 적용하여 거리오차를 최소화 하여, 최적의 위치를 찾을 수 있다.

## II. Fingerprinting 방식

Fingerprinting는 DCM 알고리즘에서 대표적인 기법이라고 할 수 있다. 경험적인 데이터에 기반한 방법으로써 먼저 위치를 측정하고자 하는 지역을 셀 형태로 나누어 각 셀에 대한 신호 값을 조사하여 저장해 놓은 다음 실제 측위 시 저장된 값과 비교하여 위치를 결정하는 방법이다. 데이터 수집단계에서는 위치 측위 대상 지역을 셀 형태로 나누어 각 셀별로 신호 강도 데이터를 기반으로 특정한 알고리즘을 사용하여 실제 위치를 결정한다. 이 방법은 특정 지역에 대한 신호 특성 값을 추출하기 때문에 주위 환경 특성을 반영한 신호 값을 저장할 수 있으며, 주위 환경이 크게 변하지 않는 한 그 지역에 대한 신호 특성 값은 대체로 일정하게 유지되기 때문에 저장된 신호 값과 현재 특정된 신호 값의 비교를 통하여 보다 정확한 위치 측위가 가능하다. 확률적 모델링에 근거한 방식은 수집된 전파 특성값이 전파 환경까지 포함되어 있기 때문에 기존의 알고리즘에 비해 우수한 위치 정확도를 제공한다. 그러나 확률적 모델링을 이용한 방식은 위치 추정을 수행하기 위하여 반드시 데이터베이스를 구축하고, 측위 정확도를 향상시키기 위하여 동일한 위치에서 여러 번의 측정 과정을 수행하여야 한다. 이러한 원리를 이용하여 측위를 하는 방식을 Fingerprint 방식이라 하며, 사전조사에 의하여 구축된 신호세기 DB와 단말에서 측정된 신호세기 값을 비교하여 위치를 결정하며 Database Correlation 방식이라 한다. 이 방식은 Database를 구축하는 단계와 측위 단계로 구성된다. 그리고 측정해야 하는 곳의 환경이 변할 때마다 새로운 데이터베이스를 다시 구축하여야 하는 문제점을 가진다[1].

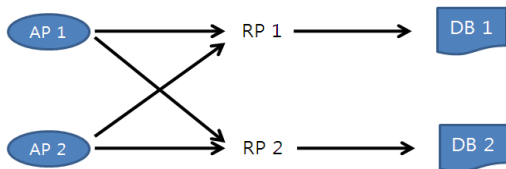


그림 1. 측위 DB생성 방법

## III. 제안 방법

### 1. KNN (K - Nearest Neighbor)

K- Nearest Neighbor(KNN)은 기억 기반 추론기법중에 하나로 패턴 인식에 많이 사용되고 있다. NN방법은 계산된 거리정보(Manhattan distance(q=1), Euclidean distance (q=2), 등) 중에서 가장 짧은 거리를 갖는 AP의 위치정보를 단말의 위치로 하는 방법이다.

### 2. Support Vector Machine

SVM의 목적은 학습 자료를 두 개의 집단으로 구분하는 함수인 초평면(Hyper-plane)을 추정하는 것이다. 이러한 평면은 무수히 존재할 수 있지만 두 집단 사이의 유클리드 거리(Margin)를 최대화 하도록 제한을 두면 하나의

유일한 평면만이 해가 된다. 이 평면을 최적의 분류초평면(Optimal Hyper-plane)이라 하고 이 평면을 중심으로 클래스를 이진 분류하게 된다[2].

두 집단의 경계선 사이의 거리(Margin)를 최대화하기 위한 라그랑지 승수  $\alpha_i$ 를 구하고 식 (1)과 식 (2)를 사용하여 초평면의 식인  $W \cdot x + b$ 를 구하며 최종적인 결정 함수인 식 (3)을 이용하여 임의의 벡터( $x$ )가 속하는 집단을 분류해 내게 된다[3].

$$W = \sum_{i=1}^N \alpha_i y_i x_i = \sum_{i \in SV} \alpha_i y_i x_i \quad (1)$$

$$y_i \left( \sum_{i \in SV} \alpha_i y_i (x_i \cdot x) + b \right) = 1 \quad (2)$$

$$f(x, \alpha^*, b^*) = \text{sign}((W^* \cdot x) + b^*) \quad (3)$$

$$= \text{sign} \left( \sum_{i \in SV} \alpha_i^* y_i (x_i \cdot x) + b^* \right)$$

비선형적인 입력데이터에 대한 분류 정확도를 높이기 위해 가우시안 커널 함수를 적용하였으며 다중 집단의 분류가 가능한 Multi-Class SVM으로 기본 알고리즘을 확장하였다[4].

식(3)의 분류기는 커널 함수를 이용하여 식 (4)의 SVM 분류기의 최종 함수이다.

$$y'(x) = \text{sign} \left( \sum_{i=1}^N \alpha_i t_i K(x_i, x_j) + b \right) \quad (4)$$

여기서  $K(x_i, x_j) = \exp \left( \frac{-\|x_i - x_j\|^2}{2\sigma^2} \right)$ 는 커널함수로

이론적으로는 가능하나 실제로는 풀기 힘든 매핑함수를 대신하여 원자료를 고차원으로 사상시켜 특징공간 내에 선형으로 분리가능한 입력자료 집합을 만들어 주는 역할을 수행한다.

### 3. KNN/SVM 혼합 알고리즘

무선 랜 실내 측위 결정을 위해 training 단계에서 여러 개의 AP로부터 수신된 SNR(Signal to Noise Ratio)값을 측정하여 Fingerprint 방식의 데이터베이스를 구축하여 측위를 위한 어떠한 장비 없이 기존에 설치된 AP를 사용하여 KNN과 SVM 군집화 기반의 혼합된 알고리즘을 무선 랜 실내 측위로 결정한다.

KNN은 Estimation 단계에서 측정된 SNR 데이터가 주어졌을 때 유사도 함수에 기반하여 Training 단계에서 Fingerprinting 방식으로 미리 저장 되어있는 RP(Reference Point)의 SNR 데이터들의 집합에서 측정된 SNR 데이터와 가장 유사한 k개의 RP의 SNR 데이터들을 선택한다.

RP와 AP의 수를 각각 m과 n으로 가정하고 측정된 SNR 데이터와 RP의 SNR 데이터들의 집합과의 유사도는 식

(5)과 같은 거리를 이용하였다

$$D_{qi} = \left( \sum_{j=1}^n |s_j - S_{ij}|^q \right)^{1/q} \quad i = 1, 2, \dots, m \quad (5)$$

$S_{ij}$ 는 training 단계에서 저장된  $i$ 번째 RP에서  $j$ 번째 AP의 SNR 데이터,  $s_j$ 는 estimation 단계에서 측정된  $j$ 번째 AP의 SNR 데이터이다.

식 (5)을 이용하여 계산된 거리 값들 중에서 가장 가까운 거리 값을  $k$ 개 만큼 선정할 수 있고 식 (6)을 이용하여 위치를 추정한다.

$$(\hat{x}, \hat{y}) = \frac{1}{K} \sum_{i=1}^K (x_i, y_i) \quad (6)$$

$K$ 는 좌표에 대한 RP수이고,  $(x_i, y_i)$ 는  $i$ 번째 RP의 좌표이고  $(\hat{x}, \hat{y})$ 는 추정된 위치 좌표이다.

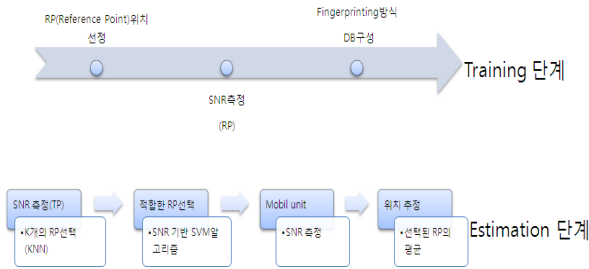


그림 2. 혼합 알고리즘 단계

위치 측위대상이 되는 공간을 일정한 범위로 나누고 각 RP(Reference Point)의 위치 값을 DB(데이터 베이스)에 저장한 후 위치 값과 AP로부터 단말기(MU: Mobile Unit)에 도달한 신호세기나 신호잡음비를 측정하여 추출한 진과 특성 값을 저장한다. 이과정은 측위대상이 되는 공간의 모든 RP를 측정할 때까지 반복적으로 수행한다.

#### IV. 모의 실험

##### 1. 실험 장비

실험 장비는 Samsung 노트북 R530을 MU로 하고, 2개의 AP를 기준으로 하여 무선 랜 신호강도를 측정하는 NetStabler 프로그램을 사용하였다. 2개의 AP로부터 각 RP포인트에 수신된 SNR 값을 연속된 그래프와 값으로 측정되었으며 1초 간격으로 SNR 값을 구하여 Fingerprinting 방식으로 데이터베이스화 하였다.

##### 2. 실험 환경

실험 장소는 (그림 2)에서 보이는 목포대학교 대외 협력관 1층 복도이다. 총길이 25m×4m인 복도에 1m 간격으로 39개의 RP 포인트와 24개의 TP 포인트를 바둑판 모양으로 지정하였고 RP 포인트는 검정색으로 TP 포인트는 빨간색으로 표시하였다. 2개의 AP로부터 1초 간격으로 RP포인트에서 100회 반복하여 측정하였고 제한한 알고리즘의 성능을 평가하기 위해 TP 포인트에서 100회 반복하여 측정하였다.

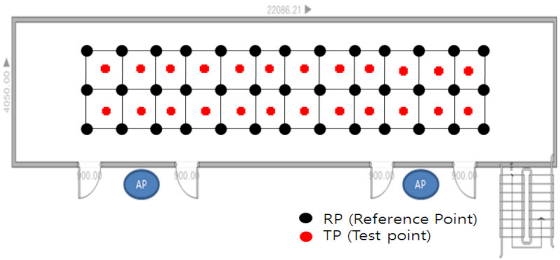


그림 3. 실험 환경 장소

##### 3. 실험 결과

식 (1)에서 모수  $k$ 와  $q$ 에 따라 KNN의 성능 결과가 다양하게 나타난다. 일반적으로 모수  $q$ 가 1 일 때는 맨하탄 거리이고 모수  $q$ 가 2 일 때는 유클리드 거리이다. (그림 4)는 알고리즘의 복잡성과 위치 예측을 고려했을 때  $q$ 는 1과 2를 사용하므로 모수  $k$ 에 따라 평균 거리 오차를 나타낸 결과이다.  $q=1$  일 때 모수  $k$ 에 따른 평균 거리 오차에서  $k$ 가 11일 때 평균 거리 오차가 1.93m로 가장 작았고  $q=2$  일 때  $k$ 에 따른 평균 거리 오차에서  $k$ 가 10 일 때 평균 거리 오차가 1.85로 가장 작았고  $q$ 가 1인 맨하탄 거리를 이용하는 KNN 방법보다  $q$ 가 2인 유클리드 거리를 이용하는 KNN 방법이 오차를 줄일 수 있다.

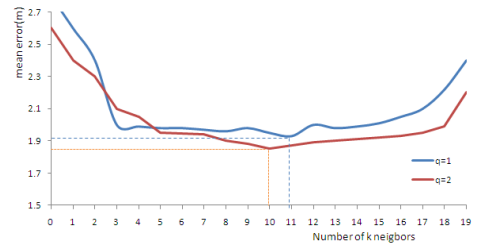


그림 4.  $k$ 개의 이웃 수와 평균 거리 오차

그림5와 그림6은 KNN과 KNN/SVM 혼합 알고리즘을 이용한 위치 추정 오차와 누적 확률에 관한 결과 이다.  $k = 10$ 일때  $q=1$ 인 맨하탄 거리와  $q=2$ 인 유클리드 거리 모두 KNN/SVM 혼합 알고리즘이 1m 이내에서는 누적확률이 비슷하지만 2-5m에서 KNN/SVM 혼합 알고리즘의 누적 확률이 즉, 정확도가 KNN보다 다소 우수하다고 할 수 있다.

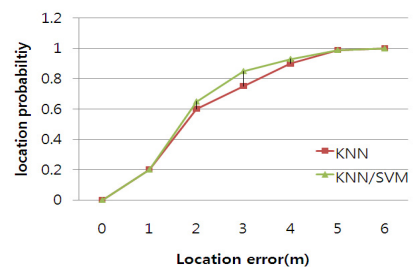


그림 5. 알고리즘 성능비교( $k=10, q=1$ )

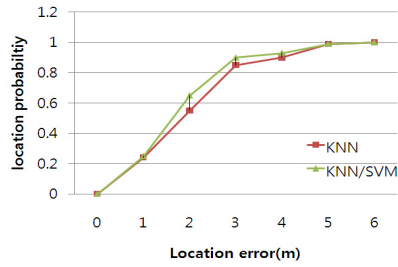


그림 6. 알고리즘 성능비교(k=10, q=2)

## V. 결론

무선 환경 하에서 패턴 매칭 알고리즘을 위해 training 단계에서 여러 개의 AP로부터 수신된 SNR값을 측정하여 Fingerprinting 방식의 데이터베이스를 만들어 활용해서 측위를 위한 특수 장비를 전혀 사용하지 않고 무선 랜을 위하여 기존에 설치된 AP를 사용하면서 KNN과 SVM 혼합된 무선 실내 측위 방법을 제안한다. KNN/SVM 알고리즘을 이용해 효율적인 실내 측위를 식별할 수 있었다.

## ACKNOWLEDGEMENT

"본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음" (NIPA-2011-C1090-1121-0007), 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구 소 지원사업으로 수행된 연구임(2011-0022980)

## 참고문헌

- [1] "WLAN 기반 실내측위 기술동향 및 현황 분석," ETRI 내부분서, 2007.5
- [2] E. Osuna, R. Freund, F. Girosi, "Support Vector Machine : Training and Applications", MIT Artificial Intelligence Laboratory and Center for Biological and Computational Learning, 1997.
- [3] M. Brunato, R. Battiti, " Statistical Learning Theory for Location Fingerprinting in Wireless LANs", Computer Networks, 2005.
- [4] J. C. Burges, " A Tutorial on Support Vector Machine for Pattern Recognition", Kluwer Academic Publishers, Boston.