

## 무선 분산 주성분 분석을 이용한 도래각 추정

김우영\*, 손경락\*, 최완°

## DoA Estimation Using Wireless Distributed PCA

Wooyoung Kim\*, Kyungrak Son\*, Wan Choi°

## 요약

고차원의 데이터를 손실을 줄이면서 저차원으로 압축시키기 위해 주성분 분석이 사용된다. 하지만 주성분 분석을 사용할 때, 규모가 큰 센서 네트워크의 데이터들을 중앙에서 처리하게 되면 통신 병목현상이 필연적으로 수반된다. 이러한 점을 개선하기 위해 탈 중앙화된 환경에서 주성분 분석을 실행하는 분산 주성분 분석에 대한 연구들이 진행되었다. 하지만, 이러한 분산 주성분 분석은 대부분 무선통신환경이 아닌, 유선통신 환경에서 진행되었다는 문제가 있다. 유선통신의 물리적 제약을 극복하고 무선 환경에 적용하기 위해 본 연구에서는 탈 중앙화된 무선통신 환경에서 분산 주성분 분석을 실행한다. 또한, 배열화된 무선 센서 노드들에서 수집한 도래각 정보를 기반으로 입사 신호의 도래각을 추정하는 방법을 제시하고 그 성능을 확인함으로써, 무선통신 환경에서 분산 컴퓨팅이 얼마나 성공적으로 동작할 수 있는지 확인해 본다.

**Key Words** : Wireless Distributed PCA, Degree of Arrival, Linear antenna arrays, Decentralized Communication, Average Consensus

## ABSTRACT

Principal component analysis is exploited to compress high-dimensional data into low-dimensions while reducing losses. However, when using principal component analysis, central processing of data from a large sensor network inevitably entails a communication bottleneck. In order to improve the communication bottleneck, studies have been conducted on the distributed principal component analysis that performs principal component analysis in a decentralized environment. However, there is a problem that most of the distributed principal component analysis was conducted in a wired communication environment rather than a wireless communication environment. To improve these physical constraints of wired communication, this study conducts distributed principal component analysis in a wireless communication environment. We estimate the direction of arrival based on the distributed principal component analysis and investigate how successfully distributed principal component analysis can operate in a wireless communication environment.

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2021R1A6A3A01086690).

※ 이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2018-0-00809, 새로운 자원을 활용한 beyond 5G 이동통신 변혁기술 개발).

♦ First Author : Samsung Research, Samsung Electronics Co. Ltd., vhehfv14@gmail.com, CL2, 정회원

° Corresponding Author : Institute of New Media and Communications and Department of Electrical and Computer Engineering, Seoul National University (SNU), wanchoi@snu.ac.kr, 중신회원

\* Institute of New Media and Communications and Department of Electrical and Computer Engineering, Seoul National University (SNU), kyungrakson@snu.ac.kr, 정회원

논문번호 : 202201-013-RN, Received January 24, 2022; Revised April 11, 2022; Accepted May 1, 2022

## I. 서론

고차원의 원본 데이터를 최대한 손실 없이 저차원으로 압축시키는데 사용되는 주성분 분석(principal component analysis)은 영상인식, 통계데이터 분석, 데이터 압축, 노이즈 제거 등 다양한 분야에 활용된다. 주성분 분석은 데이터들의 공분산 행렬이 데이터의 분산 특성을 나타낸다는 점에 기반하여 공분산 행렬의 고유값이 큰 순서대로 k개의 고유벡터를 찾아 전체 데이터를 k개의 기저를 이루는 k 차원의 데이터 공간으로 투사하여 압축한다. 이러한 주성분 기법은 대규모의 원본 데이터를 저차원의 데이터로 압축시키는데 우수한 성능을 보이지만, 주성분 분석을 사용하기 위해서 센서 네트워크에서 모은 대규모의 데이터들을 중앙으로 모아 처리하게 되면 통신 병목현상이 필연적으로 발생한다. 이러한 단점을 보완하기 위해 센서 네트워크에서 주성분 분석을 탈 중앙화된 환경에서 분산적으로 수행하는 분산 주성분 분석(distributed principal component analysis) 기법에 대한 연구가 진행되었다<sup>1-4</sup>. 이러한 센서 네트워크에서의 분산 주성분 분석 기술을 통해, 네트워크의 지능화, 특히 IoT 네트워크상에서의 지능에 대한 취합이 더욱 가속화될 것이다.

센서 네트워크와 같이, 중앙장치-연산노드 구조(master-slave architecture)가 아닌 그물망 네트워크(meshed network)에서 분산 주성분 분석을 수행하기 위해서는 average consensus(AC) 알고리즘이 필요하다<sup>5,6</sup>. AC 알고리즘이란 각 노드들의 정보에 대한 평균을 연산하는 알고리즘으로, 각 연산노드가 부분적인 데이터에 대한 연산을 수행한 후, 서로 연결된 노드들끼리 연산을 주고받으며 업데이트를 수행하여, 평균을 얻어내는 알고리즘이다. AC 알고리즘을 통해, 각 센서 노드에서 전체 데이터의 공분산 행렬을 추정하고 이를 기반으로 거듭제곱 방법(power method)을 이용하여 고유벡터의 성분을 얻는다<sup>11</sup>. 이렇게 각 노드에서 분산적으로 주성분 분석을 수행하여 각 노드가 가진 고유벡터의 성분들을 마지막에 통합하여 전체 공분산의 고유벡터를 획득한다. 하지만 분산 주성분 분석에 대한 기존연구는 유선통신 환경에서 진행되었기 때문에 실제 탈중앙화 IoT 환경에 적용하기에는 무리가 있다. 탈중앙화된 IoT 환경에 적용하기 위해서는 무선 채널을 효과적으로 처리해야 한다. 따라서 본 논문에서는 무선통신 환경에서의 분산 주성분 분석 기법을 제안한다.

본 논문에서는 무선통신 환경에서의 분산 주성분

분석 기법을 제안하기 위해, AC 알고리즘에서 각 노드 간 채널을 무선통신 채널로 모델링하고, 페이딩 현상을 제거하기 위해 다이버시티(diversity) 수준을 조정하였다. 무선 분산 주성분 분석이 정확하게 작동하는지 확인하기 위해 본 논문에서는 도래각(direction of arrival) 추정을 모의 실험하였다. 조향 벡터(steering vector)와 무선 분산 주성분 분석을 통해 얻은 고유 벡터를 이용하여 도래각을 추정하고 실제 도래각과의 차이에 기반한 오류 정도를 측정하였다. 그 후, 유선통신 환경에서 분산 주성분 분석을 통해 얻은 도래각 추정 오류와 비교하였다. 그 결과, 적정 수준의 다이버시티 수준과 SNR(signal-to-noise ratio)을 가질 때, 유선통신 환경하에서의 물리적 제약을 극복하여 무선통신 환경하에서 분산 주성분 분석이 높은 정확도로 작동함을 확인했다.

본 논문 구성은 다음과 같다. 2장에서는 주성분 분석의 배경이 되는 이론인 거듭제곱법과 AC 알고리즘에 대해 살펴보고 무선통신 모델인 레일리 페이딩 채널(Rayleigh fading channel)과 도래각 추정에 대해 소개한다. 3장에서는 레일리 페이딩 채널과 분산 주성분 분석을 결합한 무선 분산 주성분 분석을 제안하고 모의 실험 결과를 분석하였다. 마지막으로 4장에서는 본 논문의 결론을 맺는다.

## II. 본론

### 2.1 배경이론

본 절에서는 분산 주성분 분석을 위해 사용되는 기법인 거듭제곱법과 AC 알고리즘에 대해 소개한다. 주성분 분석은 데이터들의 공분산 행렬로부터 고유벡터들을 얻어내는 기법이다. 거듭제곱법과 AC 알고리즘을 이용하여 데이터들의 공분산 행렬로부터 고유벡터를 유도하는 과정을 분산적으로 진행한다.

#### 2.1.1 공분산 행렬

임의의 데이터 벡터  $\mathbf{x}$ 에 대한 공분산 행렬  $\mathbf{R}$ 의 정의는 아래와 같다.

$$\mathbf{R} = E\{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^H\} \quad (1)$$

여기서,  $(\cdot)^H$ 는 주어진 벡터의 켈레전치(conjugate transpose)를 의미한다. 실제로는 평균을 얻기 위해 총  $S$ 번의 관측에서 얻은 데이터들의 표본 공분산 행렬(sample covariance matrix)을 계산한다.

구체적으로 주어진 임의의 센서 시각  $t_s$ 에 대해,  $i$ 번째 노드에 대한 관측 데이터를  $x_i(t_s)$ , 각 노드들로부터 관측한 데이터 벡터를  $\mathbf{x}(t_s) = (x_1(t_s), \dots, x_N(t_s))^T \in \mathbb{C}^N$ 로 정의한다. 그러면, 주어진 시간  $(t_1, \dots, t_s)$  동안 관측한 데이터 벡터  $\mathbf{x}$ 에 대한 표본 공분산 행렬  $\hat{\mathbf{R}}$ 과 표본 공분산 행렬에 대한 고유값 분해는 아래와 같다.

$$\hat{\mathbf{R}} = \frac{1}{S} \sum_{i=1}^S (\mathbf{x}(t_i) - \bar{\mathbf{x}})(\mathbf{x}(t_i) - \bar{\mathbf{x}})^H = \hat{\mathbf{U}} \hat{\mathbf{\Lambda}} \hat{\mathbf{U}}^H \quad (2)$$

### 2.1.2 분산 거듭제곱법(Distributed Power Method)

표본 공분산 행렬  $\hat{\mathbf{R}}$ 의 가장 큰 고유값을 가진 고유벡터를 얻어내는 식은 아래와 같다.

$$\mathbf{v}(n+1) = \frac{\hat{\mathbf{R}}\mathbf{v}(n)}{\|\hat{\mathbf{R}}\mathbf{v}(n)\|} \quad (3)$$

여기서,  $\mathbf{v}(n) = (v_1(n), \dots, v_N(n)) \in \mathbb{C}^N$ 은 연속 분포를 가진 임의의 값들의 벡터이고,  $N$ 은 노드의 개수이다.  $\hat{\mathbf{R}}$ 의 가장 큰 고유값이 다른 고유값들 보다 더 크고,  $\mathbf{v}(0)$ 이  $\hat{\mathbf{R}}$ 의 가장 큰 고유값에 해당하는 고유벡터의 방향에 대하여 0이 아닌 성분을 가질 경우  $\mathbf{v}(n)$ 은  $\hat{\mathbf{R}}$ 의 최대 고유벡터로 수렴한다.

고유값들이 서로 다른 값이면서 양의 실수일 때,  $\hat{\mathbf{R}}$ 의  $i$ 번째 고유벡터  $\hat{\mathbf{u}}_i, i = 1, \dots, N$ 를  $\lambda_1 > \dots > \lambda_N > 0$ 에서 고유값  $\lambda_i$ 에 대응되는 고유벡터라고 정의하자. 고유벡터들은 상호 직교하므로  $\hat{\mathbf{R}}$ 의  $i$ 번째 고유벡터  $\hat{\mathbf{u}}_i$ 에 대해, 아래의 식과 같이 나타낼 수 있다.

$$(I - \hat{\mathbf{u}}_1 \hat{\mathbf{u}}_1^H) \hat{\mathbf{R}} = \sum_{i=2}^N \lambda_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^H \quad (4)$$

즉, 구한 고유벡터들을 이용해 새로운 공분산 행렬을 얻고, 그 공분산 행렬에 거듭제곱법을 사용함으로써 고유값  $\lambda_1 > \dots > \lambda_N > 0$ 의 순서대로  $N$ 개의 고유벡터들을 얻을 수 있다. AC 알고리즘을 위해 추가적인 전개를 수행하기 위해, 식 (3)에서 정규화 부분을 제거하고  $E\{\mathbf{x}\} = 0$ 이라고 했을 때 식 (2)에 의해

$$\begin{aligned} \mathbf{v}(n+1) &= \hat{\mathbf{R}}\mathbf{v}(n) \\ &= \frac{1}{S} \sum_{i=1}^S \mathbf{x}(t_i) (\mathbf{x}(t_i)^H \bullet \mathbf{v}(n)) \\ &= \frac{N}{S} \sum_{i=1}^S \left( \frac{1}{N} \sum_{j=1}^N x_j(t_i) v_j(n) \right) \mathbf{x}(t_i) \end{aligned} \quad (5)$$

라고 할 수 있다. 그러므로,

$$v_k(n+1) = \frac{N}{S} \sum_{i=1}^S \left( \frac{1}{N} \sum_{j=1}^N x_j(t_i) v_j(n) \right) x_k(t_i) \quad (6)$$

라고 할 수 있으며 특정  $k$ -노드에서  $\mathbf{x}(t_i)^H \bullet \mathbf{v}(n)$ 의 정보를 알 수 있다면 표본 공분산 행렬의 최대 고유벡터의  $k$ 번째 성분을 알 수 있다.  $\mathbf{x}(t_i)^H \bullet \mathbf{v}(n)$ 의 값을 추후 설명할 AC 알고리즘을 이용하여 이웃 노드의 정보를 반영하는 방식으로 얻을 수 있다. 마지막으로 충분히 큰  $n$ 일 때 고유벡터의 정규화를 위해 아래의 식과 같은 과정을 거치게 된다.

$$\hat{\mathbf{u}}_1 = \mathbf{v}_1(n) / \|\mathbf{v}_1(n)\| \quad (7)$$

분산 거듭제곱법에서는 각 노드는 고유벡터의 성분에 대한 정보만 가지고 있기 때문에, 정규화를 위해  $\|\mathbf{v}_1(n)\|$ 를 구하기 위해서 네트워크의 연결성을 반영한 AC 알고리즘을 이용한다.

### 2.1.3 AC 알고리즘

$n$ 개의 노드로 구성된 센서 네트워크의 상태를 나타내는 무방향 그래프(undirected graph)를 생각하자. 노드 집합  $N = \{1, \dots, n\}$ 과 변(edge)이 비순서쌍(unordered pair)  $\{i, j\} \in E$ 으로 표현되는 변의 집합  $E$ 으로 구성된 그래프를  $G = (N, E)$ 로 정의한다. 이 때, 노드  $i$ 와 연결된 노드들의 집합은  $N_i = \{j | \{i, j\} \in E\}$ 로 정의할 수 있다. 이러한 그래프 네트워크에서 노드 간의 연결을 기반으로 식 (8)과 같이 정의된 라플라시안 행렬(Laplacian matrix)  $\mathbf{L}$ 을 얻을 수 있다.

$$l_{ij} = \begin{cases} d_j & i = j \\ -1 & i \text{ communicates with } j \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

$d_j$ 는  $j$ 번째 노드와 이웃한 노드의 수를 의미한다.

AC 알고리즘은 분산 거듭제곱법을 실행하기 위해 식 (5)의  $\mathbf{x}(t_i)^H \cdot \mathbf{v}(n)$ 의 값을 얻기 위해 사용된다. 최종적으로 얻고자 하는 값  $\mathbf{x}(t_i)^H \cdot \mathbf{v}(n)$ 에 대해서 노드  $j$ 가 가진 부분적인 정보  $z_j(0, t_i)$ 를 식 (9)와 같이 정의한다.

$$z_j(0, t_i) = x_j^H(t_i)v_j(n) \quad (9)$$

정의된 식 (9)에서,

$$\mathbf{x}(t_i)^H \cdot \mathbf{v}(n) = \sum_{j=1}^N z_j(0, t_i)$$

고리즘의 가중치 파라미터  $\epsilon$ 와 식(8)과 같이 정의된 라플라시안 행렬  $\mathbf{L}$ 에 대해  $\mathbf{W} = \mathbf{I} - \epsilon\mathbf{L}$ 이라고 할 때, AC 알고리즘은 아래와 같다.

$$z_j(p+1, t_i) = W_{jj}z_j(p, t_i) + \sum_{k \in N_j} W_{jk}z_k(p, t_i) \quad (10)$$

$\bar{z}(0, t_i) = \frac{1}{N} \sum_{j=1}^N z_j(0, t_i)$ 이라고 할 때,  $\mathbf{W}$ 의 두 번째로 큰 고유값이 1보다 작고 필요충분조건이  $\mathbf{1}^T \mathbf{W} = \mathbf{1}^T, \mathbf{W} \mathbf{1} = \mathbf{1}$ 이면,

$$\frac{1}{N} \|\mathbf{z}(p, t_i) - \bar{z}(0, t_i)\mathbf{1}\|^2 \leq \lambda_2^p(\mathbf{W}) \|\mathbf{z}(0, t_i) - \bar{z}(0, t_i)\mathbf{1}\|^2 \quad (11)$$

에 의해,  $p \rightarrow \infty$ 일 때 식(9)은 찾으려는 값  $\bar{z}(0, t_i)$ 에 수렴한다. 특히,  $\lambda_2(\mathbf{W})$ 의 크기에 의해 수렴 속도가 달라지며, 충분히 큰  $p$ 에 대해서 전체 평균에 대한 추정을 만족한다<sup>11</sup>.

## 2.2 풀고자 하는 문제

배열 신호 처리 분야의 하나인 수신 신호의 도래각 추정은 신호원의 위치를 탐지하는데 있어서 중요한 과정이다. 그에 따라 도래각을 추정하는 다양한 방법에 대한 연구들이 존재한다<sup>8,9</sup>. 배열 안테나의 종류는 선형 배열, 평면 배열, 3차원 곡면 배열 안테나가 있으며, 근래에는 위상 배열안테나 또한 연구되고 있다. 무선 IoT 환경을 고려한다면, 센서의 위치가 무작위적인 상황을 고려하는 것이 가장 적합할 것이다. 하지만 본 논문에서는 우선 도래각 추정 시 필요한 분산 추정

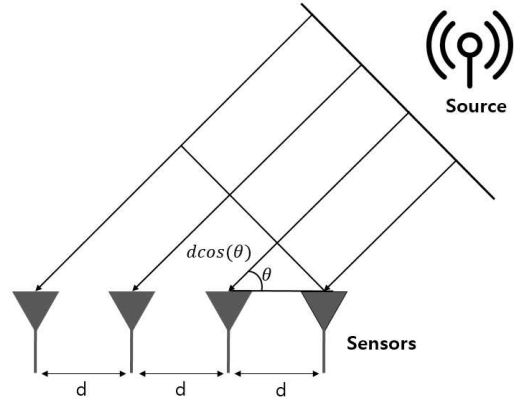


그림 1. d의 거리 만큼 떨어진 4개의 센서로 이루어진 균일 선형 배열 안테나  
Fig. 1. Uniform Linear Array composed of 4 sensors with distance d.

분 분석에 대한 연산을 분산하는 것에 집중하고자 하였다. 그래서 센서의 신호 수집은 가장 간단한 균일 선형 배열 안테나로 각 센서의 위치를 설계하여, 그림 1과 같은 균일 선형 배열 안테나(Uniform Linear Array, ULA) 배치에서 센서의 수를 20개로 증가시킨 환경을 가정한다. 그 후, 각 센서끼리 정보를 무선통신 채널을 통해 교환하면서 수신 신호들의 공분산 행렬을 얻고, 최대 고유벡터와 조향 벡터를 구한 뒤 도래각을 추정한다.

### 2.2.1 도래각 추정

그림 1처럼 구성된 균일 선형 배열 안테나에서 수신하는 신호는  $\mathbf{x}(t) = \mathbf{a}(\theta)s(t) + \mathbf{n}(t)$ 로 표현 가능하다.  $s(t)$ 는 방향을 모르는 신호원에서 나오는 신호이며  $s(t) = Ae^{j\omega t + \phi(t)}$ 이다.  $\mathbf{n}(t)$ 는 부가잡음(additive noise)이며,  $\mathbf{a}(\theta)$ 는 조향벡터로서 균일 선형 안테나 배열에서 발생하는 신호의 위상차를 나타낸다. 조향벡터는 입사각과 안테나 사이의 거리  $d$ 를 고려했을 때 식(12)로 표현할 수 있다.

$$\mathbf{a}_{ula}(\theta) = g(\theta)[1 e^{-jkd\cos(\theta)} \dots e^{-j(L-1)kd\cos(\theta)}] \quad (12)$$

식(12)와 공분산 행렬의 최대 고유벡터를 이용해서 신호원의 위치를 추정하는 식은 아래와 같다<sup>5</sup>.

$$\theta_{source} = \max_{\theta} \{ |\mathbf{a}(\theta)^H \cdot \mathbf{u}_{max}| \} \quad (13)$$

$\mathbf{a}(\theta)$ 는 조향벡터를 의미하며  $u_{\max}$ 는 수신 신호로부터 만들어진 공분산 행렬의 최대 고유벡터를 의미한다. 여러  $\theta$ 들 중  $|\mathbf{a}(\theta)^H \cdot u_{\max}|$ 를 최대로 만드는  $\theta$ 를 찾는 방식으로 도래각을 추정한다.

2.2.2 센서 노드 간 무선 채널 모델

무선통신 채널에서는 전파 경로상 건물이나 지형 등에 의한 영향으로 수신된 신호의 진폭이 변동하는 페이딩(fading) 현상이 생기게 되는데, 본 논문에서는 AC 알고리즘에서 각 노드 간 채널을 무선통신 채널 중 하나인 레일리 페이딩 채널을 적용한다. 송신 신호  $z_j(p, t_i)$ 는 레일리 페이딩 채널을 거치면서 레일리 페이딩 현상을 겪게 되고, 수신 노드에서 식(14)와 같이 변한다<sup>[10]</sup>.

$$y = h z_j(p, t_i) + w \tag{14}$$

식 (14)에서  $h$ 와  $w$ 는 각각 복소 정규분포(complex normal distributrion)를 따르는 확률 변수이다.

III. 실험

3.1 제안 기법

본 절에서는 각 센서의 수신 신호들에 대한 분산 주성분 분석을 수행하기 위하여, 2장에서 서술한 배경 이론들을 기반으로 한 무선 분산 주성분 분석의 알고리즘을 제안한다.

3.1.1 고유벡터 유도 알고리즘

식 (5)를 기반으로한 거듭제곱법 알고리즘을 통해 공분산 행렬의 고유벡터를 구할 수 있다. 공분산 행렬

의 최대 고유벡터는 식(13)에서 도래각을 추정하는데 이용된다. 거듭제곱법을 이용하여 N개의 고유벡터를 구하는 알고리즘은 아래와 같다<sup>[11]</sup>. 본 알고리즘에서 정확한 최대 고유벡터를 얻기 위해서는 각 센서가 수신한 정보들의 공분산 행렬을 정확하게 구하는 것이 중요하다. 본 논문에서는 무선통신 환경에서 AC 알고리즘을 적용하여 얻은 공분산 행렬을 기반으로 본 알고리즘을 수행한다.

3.1.2 채널 균형화(channel equalization)

페이딩으로 인한 신호 품질의 저하를 보완하기 위해, 다양한 채널 균형기법을 적용할 수 있다. 본 논문에서는 같은 정보를 시간 간격을 두고 여러 번 전송하여 서로 다른 페이딩을 겪게 한 뒤 다시 복구함으로써 깊은 페이딩(deep fading)의 영향을 줄이는 채널 균형 기법을 적용해볼 것이다<sup>[8]</sup>. 송신부에서 보낸 정보가 페이딩을 겪고 수신부에 도착하게 되면 신호는 아래와 같이 나타낼 수 있다.

$$\mathbf{y} = \mathbf{h} x_i + \mathbf{w} \tag{15}$$

식 (15)에서, 시간적으로 D번 반복 전송했을 때 수신 신호는  $\mathbf{y} = [y_1, \dots, y_D]^t$ 고 페이딩 벡터는  $\mathbf{h} = [h_1, \dots, h_D]^t$ , 부가잡음 벡터는  $\mathbf{w} = [w_1, \dots, w_D]^t$ 이다. 수신부에서 얻은 신호 벡터  $\mathbf{y}$ 에  $\frac{\mathbf{h}^H}{\|\mathbf{h}\|}$ 를 곱하게 되면 식 (16)과 같이 된다.

$$\frac{\mathbf{h}^H}{\|\mathbf{h}\|} \mathbf{y} = \|\mathbf{h}\| x_i + \frac{\mathbf{h}^H}{\|\mathbf{h}\|} \mathbf{w} \tag{16}$$

식(16)에서  $\|\mathbf{h}\|$ 를 나누게 되면, 잡음이 어느 정도 섞인 정보  $x_i$ 를 복구할 수 있다. AC 알고리즘에서 레일리 페이딩 채널을 통해 노드 간 통신을 할 때, 페이딩의 영향을 줄이기 위해서 채널 균형화가 적용된다.

3.1.3 AC 알고리즘 적용

AC 알고리즘은 이웃한 노드의 값들을 반영하면서 결국에는 전체 네트워크의 전체 평균을 추정한다. 함수의 인자(argument)는 하나의 값이거나 여러 값으로 구성된 벡터가 될 수 있으며, 가중치는 네트워크에서 가장 많은 이웃 노드를 가진 노드의 이웃 수 (maximum degree of network)의 역수인  $\frac{1}{d_{\max}}$ 로

AC Algorithm, AC(*)
1. <b>Initialization</b>
2. Set_Iteration(iteration_times);
3. <b>Message_Received:</b>
4. <b>For</b> _ = 1 : iteration_times
5. <b>For</b> i = 1 : size(neighbor_table)
6.         sum+=weight*(RF(neighbor_table[i].value)-value);
7. <b>end for</b>
8.     Send_msg(ID, value, neighbor_table);
9.     value = value + sum;
10. <b>end for</b>

Derivation of eigen vectors

1. Derivation\_first\_eigenvector

$$2. \hat{\mathbf{R}}_1 = \frac{1}{S} \sum_{i=1}^S \mathbf{x}(t_i) \mathbf{x}(t_i)^H$$

$$3. \hat{\mathbf{u}}_1 = \text{power\_method}(\hat{\mathbf{R}}_1)$$

4. Derivation\_eigenvectors

5. For i = 2 : N

$$6. \hat{\mathbf{R}}_i = (I - \hat{\mathbf{u}}_{i-1} \hat{\mathbf{u}}_{i-1}^H) \hat{\mathbf{R}}_{i-1}$$

$$7. \hat{\mathbf{u}}_i = \text{power\_method}(\hat{\mathbf{R}}_i)$$

8. end for

설정했다. 무선통신 환경에서 AC 알고리즘이 진행되기 때문에 노드 간 통신에 레일리 페이딩 채널 함수 RF(\*)를 적용한다.

3.1.4 무선 분산 주성분 분석을 이용한 도래각 추정

k-노드의 관점에서 레일리 페이딩 채널을 적용한 AC 알고리즘과 거듭제곱법을 이용해 고유벡터의 k번째 성분을 얻는다. 이렇게 각 노드들이 분산적으로 구한 고유벡터의 성분들을 모아 전체 고유벡터를 만들고 이 고유벡터를 이용하여 도래각을 추정한다.

DoA estimation (algorithm performed at node k)

1. Initialization

$$2. \mathbf{x}_k = (x_k(t_1), \dots, x_k(t_S));$$

$$3. \mathbf{v}_k = \text{rand}(S);$$

4. Power\_Method

5. For i = 1 : n

$$6. \mathbf{v}_k = N * \text{AC}(\mathbf{v}_k * \mathbf{x}_k^H);$$

$$7. \mathbf{v}_k = \text{SUM}(\mathbf{v}_k * \mathbf{x}_k) / S * \text{ones}(S)$$

8. end For

9. Normalization

$$10. u_k = \mathbf{v}_k(1)$$

$$11. c = u_k \text{conj}(u_k)$$

$$12. c = \text{sqrt}(N \text{AC}(c))$$

$$13. u_k = u_k / c$$

14. Aggregation

15. e

get\_eigenvector\_components(Network)

16. DoA\_estimation

$$17. \theta_{source} = \max_{\theta} |\mathbf{a}(\theta)^H \cdot \mathbf{e}|$$

3.2 성능 분석

분산 주성분 분석은 각 노드에서 분산적으로 연산을 진행하기 때문에 총 노드의 수는 실험 결과에 영향을 미치지 않는다. 총 노드의 수는 고정된 뒤, 다양한 실험환경에서 무선 분산 주성분 분석이 유선통신 환경에서의 분산 주성분 분석과 비교하여 얼마나 정확한지 분산 주성분 분석의 결과를 도래각 추정에 이용하여 비교했다. 유선통신 환경에서 분산 주성분 분석을 이용한 도래각 추정에 대한 결과는 선행연구<sup>11)</sup>를 참고하여 본 논문의 실험환경에 맞게 개선하여 얻었으며, 이를 본 논문에서 주장한, 무선 분산 주성분 분석을 이용한 도래각 추정 결과와 비교하였다. 유선상에서 연구된 분산 주성분 분석 대비 무선 분산 주성분 분석이 얼마나 성능이 열화되는지 확인하고, 채널 균형을 통해 성능을 개선하였다.

3.2.1 이웃노드 개수에 따른 도래각 추정 오류

SNR이 높다면 이웃 노드의 수와 관계없이 오류율이 항상 낮을 수 있다. 이웃 노드의 개수에 따른 도래각 추정 오류(Estimation Error)를 명확하게 살펴보기 위해서 SNR을 0dB로 설정하고 실험을 진행하였다.

표 1. 이웃 노드 개수에 따른 도래각 추정 실험환경  
Table 1. Simulation parameter(DoA estimation to # of neighbors).

Parameters	Value
SNR(dB)	0dB
# of sensors	20
AC Algorithm iteration	20
Power Method iteration	5
# of Neighbor nodes	2n where 0<n<9

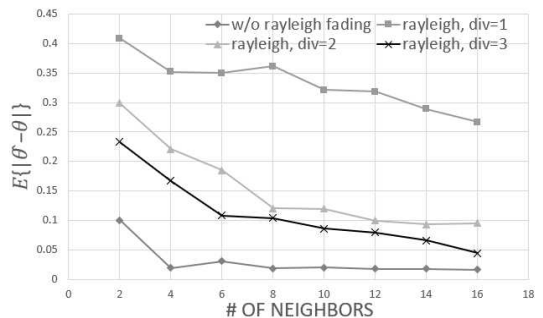


그림 2. 연결된 이웃 노드의 개수에 따른 도래각 추정 오류  
Fig. 2. DoA estimation error to # of neighbors.

그림 2에서, 각 노드마다 연결된 노드의 수가 증가할수록 추정 오류가 감소하는 모습을 보인다. AC 알고리즘에서 서로 연결되어있는 노드들이 많을수록 전체의 데이터를 반영하는 정도가 높아져, 표본 공분산 행렬의 추정에 대한 정확도가 상승하기 때문이다. 또한, 무선 채널이 적용되지 않고 유선에 의한 무오류를 가정했을 때가 추정 오류가 가장 낮지만, 레일리 페이딩 채널에서 적정 수준의 다이버시티 수준을 설정했을 때에는 유선통신에 못지않은 정확도를 보이는 것을 알 수 있다.

### 3.2.2 SNR에 따른 도래각 추정 오류

SNR이 높아질수록 노이즈 대비 신호의 강도가 강해지기 때문에, 노이즈의 영향력이 줄어들게 되고 이 역시 추정 오류를 낮추는 효과를 가져온다.

그림 3에서, SNR이 증가할수록 오류가 급격하게 줄어드는 것을 확인할 수 있다. 특히 다이버시티 수준이 2나 3일 때는 유선통신과 비교했을 때도 오류가 없다는 것을 확인할 수 있다.

표 2. SNR에 따른 도래각 추정 실험환경  
Table 2. Simulation parameter(DoA estimation to SNR).

Parameters	Value
# of Neighbor nodes	8
# of sensors	20
AC Algorithm iteration	20
Power Method iteration	5
SNR(dB)	0, 4, 8, 12, 16, 20

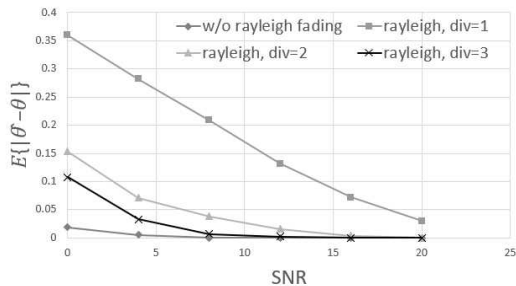


그림 3. SNR에 따른 도래각 추정 오류  
Fig. 3. DoA estimation error to SNR.

### 3.2.3 AC 알고리즘의 반복 수에 따른 도래각 추정 오류

AC 알고리즘의 반복 횟수가 증가할수록, 즉 식(11)의 가 증가할수록 표본 공분산 행렬의 추정에 대한 정

확도가 상승한다. 적절한 세기의 신호가 주어졌을 때 AC 알고리즘의 반복 횟수에 따라 무선 분산 주성분 분석 알고리즘이 유선 분산 주성분 분석과 비교해서 얼마나 열화되는지를 확인하기 위해 SNR을 20dB로 설정하고 실험을 진행하였다.

노드 간 통신 잡음의 영향을 줄일 수 있는 적절한 세기의 신호가 주어졌을 때, 식(11)에서 AC 알고리즘의 반복 횟수가 증가함에 따라 표본 공분산 행렬의 추정에 대한 오차가 0으로 수렴하게 된다. 그러므로 그림 4에서 도래각 추정 오류 역시 감소하다가 수렴하는 모습을 확인할 수 있다. 그림 2와 비교했을 때, 신호 세기를 증가시킨다면 반복전송을 이용한 채널 균형화 기법을 사용하지 않고도 유선 분산 주성분 분석과 비교해서 거의 열화되지 않는 모습을 확인할 수 있다.

표 3. AC 알고리즘의 반복 횟수에 따른 도래각 추정 실험환경  
Table3. Simulation parameter(DoA estimation to AC Algorithm iteration).

Parameters	Value
# of Neighbor nodes	4
# of sensors	20
SNR(dB)	20
Power Method iteration	5
AC Algorithm iteration	1, 2, 3, 4, 5

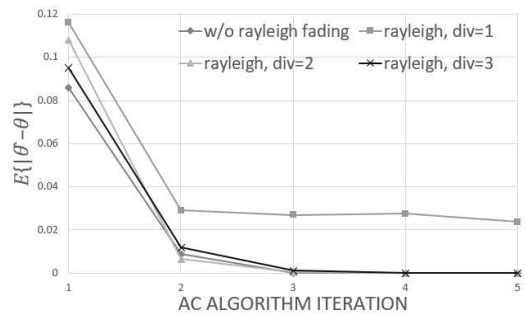


그림 4. AC 알고리즘의 반복 횟수에 따른 도래각 추정 오류  
Fig. 4. DoA estimation error to AC Algorithm iteration.

### 3.2.4 거듭제곱법 수에 따른 도래각 추정 오류

거듭제곱법의 반복 횟수를 증가시키면 표본 공분산 행렬의 최대 고유벡터 추정에 대한 정확도가 상승한다. 적절한 신호 세기가 주어졌을 때 거듭제곱법의 반복 횟수에 따라 무선 분산 주성분 분석 알고리즘이 유선 분산 주성분 분석과 비교해서 얼마나 열화되는지를 확인하기 위해 본 실험에서는 SNR을 20dB로 설

정하고 실험을 진행하였다.

그림 5에서, 거듭제곱법의 반복 횟수가 증가함에 따라 표본 공분산 행렬의 최대 고유벡터 추정이 정확해지면서 도래각 추정 오류가 감소하는 것을 확인할 수 있다. 또한, 식(3)에서 최대 고유벡터 추정 오차 역시 0으로 수렴하기 때문에<sup>7)</sup> 도래각 추정 오류의 감소 폭은 줄어든다. 그림2와 비교했을 때, 20dB의 SNR 수치에서 거듭제곱법의 반복 횟수를 적절한 수준으로 설정한다면 반복전송을 이용한 채널 균형화 기법을 사용하지 않고도 무선통신에 못지않은 정확도를 보이는 것을 확인할 수 있다.

표 4. 거듭제곱법 횟수에 따른 도래각 추정 실험환경  
Table 4. Simulation parameter(DoA estimation to Power Method iteration).

Parameters	Value
# of Neighbor nodes	4
# of sensors	20
AC Algorithm iteration	8
SNR(dB)	20
Power Method iteration	1, 2, 3, 4, 5

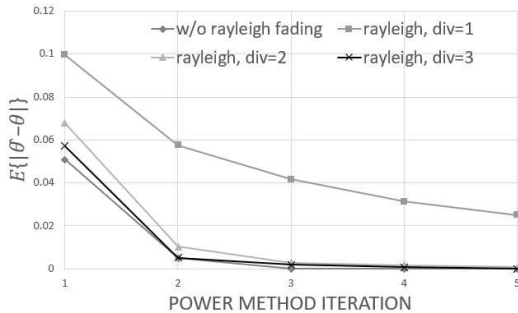


그림 5. 거듭제곱법 횟수에 따른 도래각 추정 오류  
Fig. 5. DoA estimation error to Power Method iteration.

#### IV. 결 론

SNR이 낮을 때에는 무선통신 환경에서 추정 오류가 상대적으로 높은 것을 확인할 수 있었다. 하지만, 이웃 노드의 수와 SNR, 그리고 다이버시티 수준을 적절하게 설정했을 때, 무선통신 환경에서도 오류가 거의 없이 분산 주성분 분석을 실행하여 도래각 추정이 가능함을 확인하였다.

본 연구는 각 센서의 위치를 직접 배치할 수 있다고 가정하여 가장 간단한 도래각 추정 방식인 선형 배

열 안테나를 활용하였다. 하지만, IoT 네트워크의 구조상, 분산 센서가 무작위적인 위치를 고려할 필요가 있고, 이러한 경우 각 센서에 대한 도래각이 달라지므로 문제가 더욱 복잡해지므로 이러한 문제를 해결해야 할 필요가 있어 추후 연구를 진행하고자 한다.

마지막으로, 또한 연산 능력이 있는 여러 기기들이 무선통신을 활용하고 있는 현대의 네트워크 환경상 유선통신 환경에서 진행된 분산 주성분 분석에 대한 기존 연구는 실제 탈중앙화 IoT 환경에 적용하기에 무리가 있었다. 본 연구는 DoA 기법에 대해서만 다루었으나, 앞으로 다른 연산 기법들과 센서 네트워크의 무선 분산 주성분 분석 기술 및 전송 기술을 융합하여, 네트워크의 지능화 및 데이터 처리, 특히 IoT 네트워크상에서의 지능에 대한 취합을 가속화 시키는 기술을 개발하고자 한다.

#### References

- [1] A. Scaglione, R. Pagliari, and H. Krim, "The decentralized estimation of the sample covariance," *42nd Asilomar Conf. Sign., Systems and Computers*, pp. 1722-1726, Asilomar, USA, Nov. 2008.
- [2] H. S. Won, S. P. Kim, Lee, S. H. Choi, M. J. Moon, and S. Yang, "Secure principal component analysis in multiple distributed nodes," *Secur. and Commun. Netw.*, vol. 9, no. 14, pp. 2348-2358, Sep. 2016.
- [3] S. X. Wu, H. T. Wai, L. Li, and A. Scaglione, "A review of distributed algorithms for principal component analysis," in *Proc. IEEE*, vol. 106, no. 8, pp. 1321-1340, Aug. 2018.
- [4] K. Son, A. Ramamoorthy, and W. Choi, "Distributed matrix multiplication using group algebra for on-device edge computing," *IEEE Sig. Proc. Lett.*, vol. 28, pp. 2097-2101, Oct. 2021.
- [5] S. B. L. Xiao and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," *IPSN 2005. Fourth Int. Symp. Inf. Process. in Sensor Netw.*, pp. 63-70, Los Angeles, USA, Apr. 2005.
- [6] L. Xiao, S. Boyd, and S. J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel and Distrib. Comput.*,



vol. 67, no. 1, pp. 33-46, Jan. 2007.

- [7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th Ed., The Johns Hopkins Univ. Press, pp. 330-332, 2013.
- [8] H. M. Noh, H. J. Kim, and H. N. Kim, "DOA estimation of combining time modulated arrays with beamspace MUSIC algorithm," *J. KICS*, vol. 45, no. 6, pp. 951-960, Jun. 2020.
- [9] P. Stoica, B. Ottersten, M. Viberg, and R. L. Moses, "Maximum likelihood array processing for stochastic coherent sources," *IEEE Trans. Sig. Proc.*, vol. 44, no. 1, pp. 96-105, Jan. 1996.
- [10] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge Univ. Press, pp. 76-80, 2004.

**김 우 영 (Wooyoung Kim)**



2022년 2월 : 서울대학교 전기정보공학부 학사  
2022년 1월~현재 : 삼성전자 삼성리서치 CL2  
<관심분야> 데이터 신호처리, 분산 컴퓨팅  
[ORCID:0000-0001-7130-2266]

**손 경 락 (Kyungrak Son)**



2014년 2월 : 한국과학기술원 (KAIST) 전기 및 전자공학과 학사 (수리과학과 복수전공)  
2016년 2월 : 한국과학기술원 (KAIST) 전기 및 전자공학부 석사  
2016년 3월~2021년 8월 : 한국과학기술원 (KAIST) 전기 및 전자공학부 박사과정  
2021년 9월~현재 : 서울대학교 뉴미디어통신공동연구소 선임연구원  
<관심분야> 무선통신, 정보이론, 부호이론  
[ORCID:0000-0002-2983-985X]

**최 완 (Wan Choi)**



1996년 2월 : 서울대학교 전자공학 학사  
1998년 2월 : 서울대학교 전자공학 석사  
2006년 12월 : University of Texas at Austin, Department of Electrical and Computer Engineering 박사  
2007년 1월~2020년 2월 : 한국과학기술원 (KAIST) 전기 및 전자공학부 조교수, 부교수, 정교수  
2020년 3월~현재 : 서울대학교 뉴미디어통신공동연구소 및 전기정보공학부 정교수  
<관심분야> 무선통신, 정보이론, 머신러닝  
[ORCID:0000-0003-3930-7088]