Hybrid Maximum Depth-*k*NN Method for Real Time Node Tracking using Multi-Sensor Data

Sudhir Kumar, Abhay Kumar, Akshay Kumar and Rajesh M. Hegde
Department of Electrical Engineering
Indian Institute of Technology, Kanpur, India
Email: {sudhirkr, abhayk, akshakr, rhegde}@iitk.ac.in

***Abstract*—In this paper, a hybrid maximum depth - *k* Nearest Neighbour (hybrid MD-*k*NN) method for real time sensor node tracking and localization is proposed. The method combines two individual location hypothesis functions obtained from generalized maximum depth and generalized *k*NN methods. The individual location hypothesis functions are themselves obtained from multiple sensors measuring visible light, humidity, temperature, acoustics, and link quality. The hybrid MD-*k*NN method therefore combines the lower computational power of maximum depth and outlier rejection ability of *k*NN method to realize a robust real time tracking method. Additionally, this method does not require the assumption of an underlying distribution under non-line-of-sight (NLOS) conditions. Additional novelty of this method is the utilization of multivariate data obtained from multiple sensors which has hitherto not been used. The affine invariance property of the hybrid MD-*k*NN method is proved and its robustness is illustrated in the context of node localization. Experimental results on the Intel Berkeley research data set indicates reasonable improvements over conventional methods available in literature.**

I. Introduction

Recent technological developments in microelectromechanical systems (MEMS) and wireless technology have accelerated the current research in ad-hoc sensor networks (AHSN) [1]. The basic aspects of AHSN are data acquisition, processing and decision making based on the acquired spatio-temporal information. Localization is a crucial aspect of AHSN, where routing and subsequent evaluation of path cost for the network are carried out in many applications like emergency response. Localization is a method to compute the position of the node in a 2D/3D space, while tracking is a continuous localization of the mobile nodes over time [2]. An energy-efficient and real-time target tracking sensor network is proposed in [3]. In this paper, sensor node localization [4] using data depth is described. Estimation of node location is based on proximity. The term proximity refers to the fact that predicted node location is assigned to the nearest grid point in the network. Proximity based localization requires lower computational time and provides higher localization accuracy. Some applications may not require accurate position of the sensor node, for instance, predicting the location of fire-fighters in an indoor building. The multi-sensor data used in this work are visible light, temperature wave-front, humidity, acoustic signal and link quality for locating the sensor nodes. Thus, each of the observation corresponds to a five-dimensional vector, which represents a multivariate observation. The online multivariate observation vector recorded by each of the sensor node is logged on to the central system. Using this observation vector, position of each of the sensor node can be predicted over time.

This work was supported in part by Indian Space Research Organisation (ISRO). The author S. Kumar was supported by TCS Research Scholarship Program TCS/CS/2011191C.

A. Motivation and Contributions

Univariate data is widely analysed using various statistical moments, like location, scale, skewness, and kurtosis. The localization algorithm based on inter-nodal distances assuming univariate data distribution is described in [5], [6]. Distance estimation using received signal strength is erroneous under NLOS conditions, which leads to higher localization errors. The statistical moments of univariate case can be easily extended to determine the characteristics of multivariate data [7]. However, these moments are difficult to compute for multivariate data. In many cases these moments do not exist, making this approach inappropriate. Generally, we analyse a multivariate data on the presumption of a normal distribution. But it is not always true in non-line of sight conditions. Data depth is an important technique for non-parametric analysis and inference from multivariate data [8]. It facilitates systematic way of ordering the multivariate data, this is known as center-outward ordering [9] or outlyingness [7]. Contours based on data depth analysis provide more intuitive visualisation of distributional properties as compared with statistical moments. Utilization of multi-sensor data [10] like visible light, temperature wave front, humidity, acoustic signal and link quality makes the localization algorithm efficient for the real-time applications. Experimental nodes are generally capable of measuring all these modalities simultaneously. Hence, it improves the localization accuracy.

The contributions of the paper are enumerated herein. Localization of sensor nodes using depth as a measure without any assumption of the distribution of multivariate data is proposed in this paper. The multi-sensor data collected from five different types of sensors are considered simultaneously for localization. The proposed tracking method is developed for the multiple hypothesis testing framework. Multiple hypothesis correspond to the multiple grids in the network. In this paper, a hybrid maximum depth - *k* Nearest Neighbour (hybrid MD-*k*NN) method for real time node tracking is proposed. The method generalizes the maximum depth [7], [11] and *k*NN [12] method to account for a vector of multivariate observation under multiple hypothesis. Generalized maximum depth method requires lesser computation time than the generalized *k*NN method, whereas generalized *k*NN method is robust in the non-line of sight conditions. The proposed method has higher localization accuracy than the generalized maximum

978-1-4673-6432-4/15/$31.00 ©2015 IEEE

depth method. Proposed hybrid MD-kNN method is superior to the generalized kNN method in terms of time complexity.

The rest of the paper is organized as follows. Section II presents the notations and known definitions. In Section III, hybrid MD-kNN method for sensor node tracking is described. Performance evaluation is illustrated in Section IV. Finally, a brief conclusion is presented in Section V.

II. Notations and Known Definitions

The definitions and desirable properties of different depth functions are described first. Subsequently, these depth functions are utilized in describing the hybrid MD-kNN method for node tracking. The depth functions are judiciously chosen based on the application-specific requirements like computation time and robustness of localization.

The localization of sensor nodes are carried out in a q dimensional coordinate system. Let L distinct grid points in the network are denoted by $\{G_l \triangleq (\alpha_l^1, \alpha_l^2, \ldots, \alpha_l^q)\}_{l=0}^{L-1}$. At each of the grid, G_l, n_l number of training observations points are collected during offline phase. Suppose there are N training observation points, $\mathbf{O} = \{X_1, X_2, \ldots X_N\}$. These observations are categorised into L grids namely, $G_0, G_1, \ldots, G_{L-1}$ containing $n_0, n_1, \ldots, n_{L-1}$ points respectively such that $\sum_{l=0}^{L-1} n_l = N$. Given an observation vector, $X = \{x_1, x_2, \ldots, x_m, \ldots, x_M\}$, where $x_m \in \mathbf{R}^5$, it is localized to one of the grid point using hybrid MD-kNN method. This is a function $l : X \to Y$ where, $Y \in \{0, 1, \ldots, L-1\}$ that associates an observation vector X with its corresponding nearest grid Y. In this paper, indicator function is denoted by $\mathbb{I}$.

A. Statistical Properties of Depth Functions

A random vector Z in $\mathbf{R}^d$ is said to be symmetric [9] with respect to the following types of symmetricity.

- Centrally symmetric: About θ if $Z - \theta \overset{d}{=} \theta - Z$, where $\overset{d}{=}$ denotes equal in distribution.
- Angularly symmetric: About θ if $\frac{(Z-\theta)}{\|Z-\theta\|} \overset{d}{=} \frac{(\theta-Z)}{\|Z-\theta\|}$.
- Halfspace symmetric: About θ if $P(Z \in H \geq 1/2)$ for every closed halfspace containing θ.

Depth functions are used for center outward ordering of points in $\mathbf{R}^d$. It should have following desirable properties [9]. Let D(.;.): $\mathbf{R}^d \times \mathscr{F} \to \mathbf{R}^1$ be mapping of depth of points in $\mathbf{R}^d$, whose distribution is given by $\mathscr{F}$.

1) Affine Invariance: Under affine transformation of coordinate system, the depth of a point $x \in \mathbf{R}^d$ should remain unchanged as affine transformation preserves collinearity and ratio of distances. Let the transformation be represented as $x \mapsto Ax + b$, where A is an invertible $d \times d$ matrix and b is a d-dimensional column vector, then

$$D(Ax + b; F_{Ax+b}) = D(x; F_x) \tag{1}$$

2) Maximality at Center: Any depth function should attain its maximum at center of the distribution. The center is defined as a point of symmetry with respect to some notion of symmetry.

$$D(\phi; F) = \sup_{x \in \mathbf{R}^d} D(x; F) \tag{2}$$

It holds for any distribution $F \in \mathscr{F}$ having its center at ϕ.

3) Monotonous with respect to the Deepest Point: As a point x moves away from the center along a ray passing through the center, the value of depth function should decrease monotonically.

$$D(x; F) \leq D(\phi + \alpha(x - \phi); F) \qquad \text{holds for } \alpha \in [0, 1] \tag{3}$$

4) Vanishes at Infinity: The depth function should approach zero as Euclidean distance of the point from the center approaches infinity.

$$D(x; F) = 0 \qquad \text{as } \|x\| \to \infty \tag{4}$$

B. Definition of Depth Functions

In this Section, different types of depth function are discussed.

1) Location Depth [13]–[15]: For an observation, $x \in \mathbf{R}^d$ with respect to a distribution having probability measure P on $\mathbf{R}^d$, location depth is defined as the minimum of probability mass of any closed halfspace containing x.

$$\mathcal{L}(x; P) = \inf_H \{P(H) : H \text{ is a closed halfspace, } x \in H\} \tag{5}$$

The location depth function satisfies all the four desirable properties.

2) Simplicial Depth [16]: For an observation, $x \in \mathbf{R}^d$ relative to a probability measure P on $\mathbf{R}^d$, simplicial depth is defined as the probability that x belongs to a random simplex in $\mathbf{R}^d$.

$$\mathcal{S}(x; P) = P(x \in S[X_1, X_2, X_3, ..., X_{d+1}]) \tag{6}$$

Simplicial depth functions may fail to satisfy the monotonicity property and maximality property for centrally symmetric and halfspace symmetric discrete distributions respectively.

3) Mahalanobis Depth [9], [17]: The Mahalanobis distance of a point x from $\mathbf{O}$ with mean μ and covariance matrix Σ is defined as

$$\mathcal{D}_m(x) = \sqrt{(x - \mu)^\top \Sigma^{-1} (x - \mu)} \tag{7}$$

Assuming that second moment of $\mathbf{O}$ exists, Mahalanobis depth of x_i is defined as

$$\mathcal{M}_i = [1 + \underbrace{(x_i - \bar{X})^\top \Sigma^{-1} (x_i - \bar{X})}_{\mathcal{D}_m^2(x_i)}]^{-1} \tag{8}$$

where, $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ and $\Sigma = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})^\top$. Mahalanobis Depth function may fail to satisfy maximality property for angularly symmetric distributions.

4) L_1 Depth [13], [18]: For an observation, x, with respect to $\mathbf{O}$ in $\mathbf{R}^d$, L_1 depth is defined as one minus average of the unit vectors, directed from x towards all observations points.

$$L_1(x, \mathbf{O}) = 1 - \left\| \frac{1}{N} \sum_{i=1}^N u_i(x) \right\| \tag{9}$$

where, $u_i(x) = \frac{x - X_i}{\|x - X_i\|}$ denotes the i^{th} unit vector. For x lying far away from the center, all unit vectors gets summed up i.e., $\lim_{\|x\| \to \infty} \|\bar{u}(x)\| = 1$ and hence L_1 depth approaches zero.

$$\lim_{\|x\| \to \infty} L_1(x, \mathbf{O}) = 0 \tag{10}$$

For points near center, value of L_1 depth is higher as unit vectors cancel each other.

$$0 \leq L_1(x, \mathbf{O}) \leq 1 \tag{11}$$

L_1 depth function violates affine invariance property [18].

5) Oja Depth [7]: For an observation, $x \in \mathbf{R}^d$ with respect to $\mathbf{O}$, whose distribution is given by F, Oja depth is defined as

$$\mathcal{O}(x; F) = [1 + E_F[v(S[x, X_1, \ldots, X_d])]]^{-1} \tag{12}$$

where, $S[x, X_1, \ldots, X_d]$ represents the closed simplex with vertices x and random d points from the $\mathbf{O}$. Expected volume with respect to distribution, F, is represented by $E_F[v]$. Oja depth function may violate affine invariance property.

III. Hybrid MD-kNN Method for Sensor Node Tracking

In this Section, hybrid MD-kNN method for sensor node tracking is described. The method combines two individual location hypothesis functions (LHFs) obtained from generalized maximum depth and generalized kNN methods. Subsequently, statistical properties and the algorithmic description of the proposed method is also provided.

A. Hybrid MD-kNN Method

The hybrid MD-kNN algorithm comprises of two location hypothesis functions (LHF). Generalized maximum depth and generalized kNN method are utilized to describe these LHFs. Unlike maximum depth and kNN method, generalized methods take into account a vector of multivariate observation under multiple hypothesis (multiple grid points) testing framework. The proposed hybrid MD-kNN method optimizes both computational time and robustness. It may be noted that the generalized maximum depth method has lower computation time as compared to the generalized kNN method. On the contrary, generalized kNN method is highly robust as compared to the generalized maximum depth method. Therefore, hybrid MD-kNN method has lower time complexity than the generalized kNN method, while more robust than the generalized maximum depth method.

1) Generalized Maximum Depth Method: Generalized maximum depth method utilized in the LHF is described as follows. Consider two grids namely grid 0, G_0, and grid 1, G_1, and an observation vector, $X = \{x_1, x_2, .., x_M\}$, having M points. Then the observation vector is said to be belonging to G_1 if the number of points in X, having higher depth relative to G_1 than that to G_0, is greater than the number of points in X, having higher depth relative to G_0 than that to G_1. The binary hypothesis testing problem for node tracking is given as

$$\hat{h}_D(X; G_0, G_1) = \mathbb{I}\Bigg[\overbrace{\sum_{i=1}^{M} \mathbb{I}\,[D(x_i, G_1) > D(x_i, G_0)]}^{n(X,G_1)} > \underbrace{\sum_{i=1}^{M} \mathbb{I}\,[D(x_i, G_0) > D(x_i, G_1)]}_{n(X,G_0)} \Bigg] \tag{13}$$

where, $n(X, G_l)$ represents the number of points in observation vector X having higher depth with respect to G_l $\forall$ $l = \{0, 1\}$. Generalized maximum depth method for a multivariate observation vector for binary hypothesis problem can be generalised for multiple hypothesis framework as following

$$\hat{h}_D(X; G_0, G_1, \ldots, G_{L-1}) = \operatorname*{argmax}_l \Big[n(X, G_l) \Big] \tag{14}$$

If maximum number of points of online observation vector belongs to G_l then the observation vector is localized to a position corresponding to G_l in the network grid.

2) Generalized kNN Method: 'Maximality at center' property of depth function and symmetrization with respect to x are exploited to define generalized kNN. Subsequently, an x-outward ordering of points is constructed. Symmetrization construction involves adding to the training observation points $\mathbf{O} = \{X_1, X_2, \ldots, X_N\}$, their reflections with respect to x (i.e., $2x - X_1, 2x - X_2, .., 2x - X_N$). As a result of symmetrization, x becomes the deepest point of the final training points. Generalized kNN method is used to localize an online observation vector, $X = \{x_1, x_2, \ldots, x_M\}$, having M points. Let $\mathbf{R}_x^\beta$ denote the smallest depth-based neighbourhood [12] that contains atleast β proportion of total training points, where $\beta = \frac{k}{N}$. $\hat{m}_D^\beta(x_j)$ will be 1 if j^{th} point of observation vector belongs to G_1 and lies in $\mathbf{R}_x^\beta$. Hence, $\sum_{j=1}^{M} \hat{m}_D^\beta(x_j)$ signifies the total number of points of the observation vector belonging to G_1 in $\mathbf{R}_x^\beta$. Similarly $1 - \hat{m}_D^\beta(x_j)$ will be 1 if x_j belongs to G_0 in $\mathbf{R}_x^\beta$ and hence, $\sum_{j=1}^{M}[1 - \hat{m}_D^\beta(x_j)]$ signifies the total number of points of the observation vector belonging to G_0 in $\mathbf{R}_x^\beta$. If the total number of points of observation vector belonging G_1 in $\mathbf{R}_x^\beta$ is more than those belonging to G_0 in $\mathbf{R}_x^\beta$, the observation vector is localized to G_1 and vice versa.

$$\hat{m}_D^\beta(X; G_0, G_1) = \mathbb{I}\Bigg[\sum_{j=1}^{M} \overbrace{\mathbb{I}\Big(\sum_{i=1}^{N} \mathbb{I}\,[Y_i = 1]\, W_i^\beta(x) > \sum_{i=1}^{N} \mathbb{I}\,[Y_i = 0]\, W_i^\beta(x) \Big)}^{\hat{m}_D^\beta(x_j)} > \sum_{j=1}^{M} \underbrace{\mathbb{I}\Big(\sum_{i=1}^{N} \mathbb{I}\,[Y_i = 0]\, W_i^\beta(x) > \sum_{i=1}^{N} \mathbb{I}\,[Y_i = 1]\, W_i^\beta(x) \Big)}_{1 - \hat{m}_D^\beta(x_j)} \Bigg] \tag{15}$$

with $W_i^\beta(x) = \frac{1}{K_x^\beta} \mathbb{I}\Big[X_i \in \mathbf{R}_x^\beta\Big]$ where $K_x^\beta = \sum_{j=1}^{N} \mathbb{I}\Big[X_j \in \mathbf{R}_x^\beta\Big]$ denotes total number of points in $\mathbf{R}_x^\beta$. Let us consider a multi-grid scenario for the given training points, $\{X_1, X_2, \ldots, X_N\}$, where an observation vector, $X = \{x_1, x_2, \ldots, x_M\}$, is localized to any one of the L grids, namely, $\{G_0, G_1, \ldots, G_{L-1}\}$ using kNN method. The generalized kNN method for multiple hypothesis testing scenario is formulated as

$$\hat{m}_D^\beta(X; G_0, G_1, ..., G_{L-1}) = \operatorname*{argmax}_l \Big[\sum_{j=1}^{M} \mathbb{I}\Big(\sum_{i=1}^{N} \mathbb{I}[Y_i = l] W_i^\beta(x) \Big) \Big] \tag{16}$$

where $W_i^\beta(x)$ is as defined in Equation 15. The l^{th} grid, G_l, is chosen corresponding to the maximum argument.

3) Hybrid MD-kNN Method using LHFs: The expression for hybrid MD-kNN tracking method is given as the sum of two LHFs. In this method, threshold (τ) is a value that affects localization robustness and time complexity of sensor nodes. As τ increases, localization robustness and time complexity increases and vice-versa. For observation points having depth greater than $\tau * MD$, maximum depth method is incorporated. And for observation points having depth upto $\tau * MD$, kNN method is used.

$$\hat{H}_D(X; G_0, \ldots, G_{L-1}) = \overbrace{\hat{h}_D(X; G_0, \ldots, G_{L-1})\mathbb{I}\Big[D(X) > \tau * MD\Big]}^{LHF_1} + \underbrace{\hat{m}_D^{\beta(n)}(X; G_0, \ldots, G_{L-1})\mathbb{I}\Big[D(X) \le \tau * MD\Big]}_{LHF_2} \quad (17)$$

Where MD denotes the maximum depth value obtained from the observation collected at grid points, $G_l \ \forall \ l$. The threshold can be reduced to optimize the computation time, at the expense of robustness. Similarly increasing the threshold improves robustness, but increases computation time. Hence, there is a trade-off between computation time and localization robustness.

B. Statistical Properties of the Hybrid MD-kNN Method

1. Affine Invariance: Aforementioned hybrid MD-kNN method is defined as a function of various depth functions, like location depth, simplicial depth, and Mahalanobis depth. Since these depth functions are affine invariant, hybrid MD-kNN method would also be affine invariant.

Proof:
Let $X \mapsto Y$ and $Y = AX + b$ represent the affine transformation. $\hat{h}_D(x; G_{1x}, G_{0x})$ denote the maximum depth method for original observation and $\hat{h}_D(y; G_{1y}, G_{0y})$ denote the maximum depth method for the affine-transformed observation.

$$\begin{aligned} \hat{h}_D(y; G_{1y}, G_{0y}) &= \mathbb{I}\Big[D(y, G_{1y}) > D(y, G_{0y})\Big] \\ &= \mathbb{I}\left[D(x, G_{1x}) > D(x, G_{0x})\right] \\ &= \hat{h}_D(x; G_{1x}, G_{0x}) \end{aligned} \quad (18)$$

The first and third equalities are induced from the definition of maximum depth method. The second equality follows if $D(.;.)$ is affine-invariant. Hence, maximum depth method for the observation vector is affine invariant. It may be noted that the kNN method is affine-invariant [12]. Therefore, hybrid MD-kNN method is also affine-invariant.

2. Robustness Analysis: Most statistical depth functions will assign almost zero depth value to an outlier point x with respect to any grids, G_l. making it impossible to be localized using generalized maximum depth method.

$$\lim_{\|x\| \to \infty} D(x, G_l) = 0 \quad (19)$$

Symmetrization construction with respect to x, involved in defining depth-based neighbourhood, makes x the center of the resulting symmetrized training observations. Now, its depth value is non-zero and can be easily localized using generalized kNN depth method. Since kNN depth method is based on depth-based neighbourhood, it is robust to outliers [12]. Therefore, hybrid MD-kNN method is robust than the maximum depth method.

C. Algorithm for Hybrid MD-kNN for Sensor Node Tracking

In this Section, algorithm for hybrid MD-kNN method for tracking is summarized in Algorithm 1.

Algorithm 1 : Hybrid MD-kNN Method for Node Tracking

1: **Input** : Grid points, $\{G_l \triangleq (\alpha_l^1, \alpha_l^2, \ldots, \alpha_l^q)\}_{l=0}^{L-1}$. Choose the threshold τ according to desired localization accuracy and computation time.
2: **Measurement Acquisition** : Each element of observation vector is a multivariate observation containing visible light, temperature wave-front, humidity, acoustic signal and link quality as labels.
3: **Offline Training** : Multivariate observation vectors are recorded at each grid point. At each of the grid, G_l, n_l training points are collected during offline phase.
4: **Iteration** : Repeat for all the nodes at each time instant.
5: **Online Testing** : Record online observation vector, $X = \{x_1, x_2, .., x_M\}$, where $x_m \in \mathbf{R}^5$, at time t corresponding to a particular sensor node.
6: **Node Localization** : Localize to one of the nearest grid point using the hybrid MD-kNN method in Equation 17.
7: **Condition check** : Are all the nodes localized, if not, go to Step 4.
8: **Termination** : Output, the location of all nodes.

IV. Performance Evaluation

In this Section, experimental conditions are described first. Subsequently, experimental results for sensor node tracking using Intel Berkeley research lab database [19]. Experimental results from real field deployment are also reported.

A. Experimental Conditions

The sensor nodes (Mica2Dot) considered for the Intel Berkeley research lab database are shown in Figure 1. Fifty four nodes are randomly deployed in $40m \times 30m$ indoor scenario. TinyOS platform are used for collecting the Intel data using the TinyDB in-network query processing system [19]. Crossbow motes and MTS310 sensor boards are deployed as sensor nodes in Figure 2. XM2110 IRIS board and MIB520 USB mote interface are used as a gateway to configure the network. Also, cross-validation of the performance of the proposed algorithms is carried out using National Instruments (NI) 9792 gateway, WSN node - 3212, 3202. The communication among sensor nodes are based on IEEE 802.15.4 protocol.

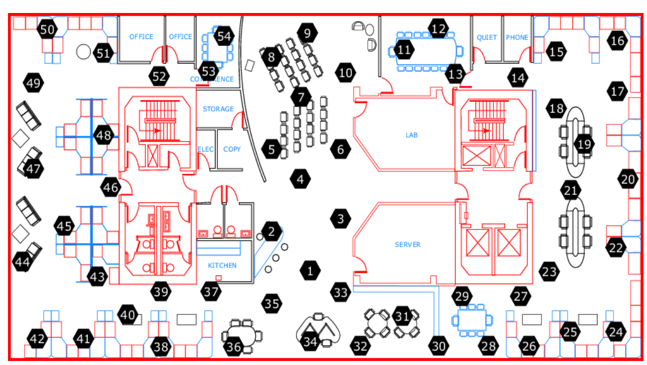

Fig. 1. Figure illustrating the deployments of fifty four sensor nodes in Intel Berkeley Research lab. Reproduced from [19].

For experimental deployment, shoe-mounted sensor node is utilized for locating the fire-fighter in an indoor scenario as shown in Figure 2. Each of the fire-fighter acts as a mobile sensor node. Initially, training is performed coarsely to construct

TABLE I. COMPARISON OF LOCALIZATION ACCURACY AND PROBABILITY OF RESOLUTION FOR VARIOUS METHODS USING DIFFERENT DEPTH FUNCTIONS

	Localization Accuracy (m)			Probability of Resolution (POR)		
Depth Functions	**Maximum Depth**	**Hybrid MD-*k*NN (τ = 0.2)**	***k*NN**	**Maximum Depth**	**Hybrid MD-*k*NN (τ = 0.2)**	***k*NN**
Location Depth	1.1598	0.8805	0.5445	0.83	0.87	0.91
Simplicial Depth	0.8475	0.7556	0.5245	0.88	0.89	0.90
Mahalanobis Depth	1.4969	1.3014	0.5794	0.78	0.80	0.89
Oja Depth	2.1092	2.1092	0.5840	0.69	0.69	0.90
L_1 Depth	2.7202	1.8094	0.6370	0.63	0.75	0.89

a raw data map of each grid point. Grid point herein refers to the center of each room of the building. However, more grid points can be taken into account to achieve higher localization accuracy, depending upon application specific requirements.

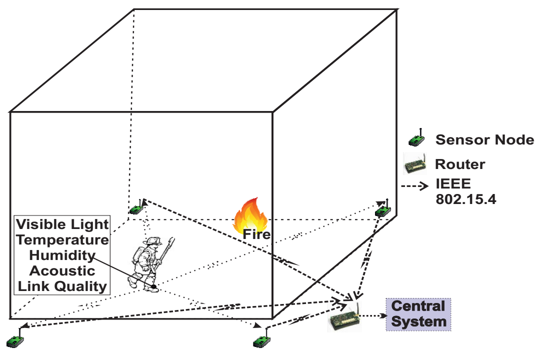

Fig. 2. Figure illustrating the experimental set-up for firefighter tracking in indoor scenario.

B. Experimental Results

Performance metrics used for evaluating the performance of proposed algorithm are localization accuracy, probability of resolution and time complexity. Localization error for a node is defined as the Euclidean distance between actual and predicted node location. Probability of resolution (POR) is defined as the proportion of sensor nodes allocated to the correct grid points. The localization is carried out on a PC with i5-3330 processor @2.6 GHz and 4 GB RAM.

TABLE II. TABLE ILLUSTRATING THE COMPARISON OF COMPUTATION TIME AND LOCALIZATION ACCURACY FOR VARIOUS METHODS

Methods	Computation Time	Robustness of Localization
Maximum depth-based	Small	Average
***k*NN depth-based**	Longer	Better
Hybrid MD-*K*NN	Average	Good

1) Time Complexity Analysis for Hybrid MD-kNN Method: The optimal algorithm to compute the location depth of a point, X, in $\mathbf{R}^2$ is $O(NlogN)$ [20]. Simplicial depth can be computed in $O(N^2)$ [21]. The optimal algorithm to compute Oja depth has $O(Nlog^3N)$ time complexity [20]. L_1 depth and Mahalanobis depth have $O(N)$ computation complexity. There is a trade-off between localization accuracy and time complexity. Localization using simplicial and location depth functions have highest time complexity as depicted in Figure 3. However, at the same time, method using these depth functions have lower localization error. The computation time shown here is in fraction of seconds, which provides real-time solution of node tracking and related applications. Hence, τ is judiciously chosen based on application specific requirements.

For a large number of training observation points, the expected volume of all simplices does not vary much for different elements of the observation vector. From the definition of Oja depth, it follows that depth values has narrow range. In hybrid MD-*k*NN method, *k*NN method is incorporated for points, which have depth value less than $\tau * MD$. For lower value of τ, none of the points satisfy this criteria . In fact, all the points are localized using maximum depth method. Therefore, time complexity for Oja depth is minimal. Note that, it is found that *k*NN method becomes effective only after $\tau = 0.96$ for this scenario, which results in high time complexity at $\tau = 1$, for Oja depth.

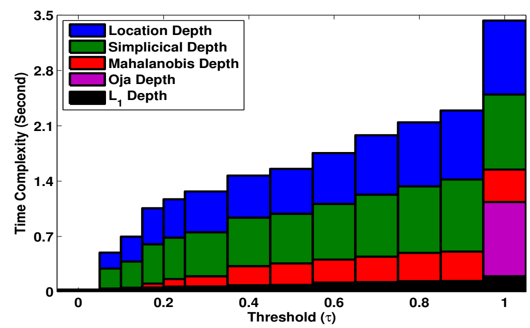

Fig. 3. Figure illustrating the time complexity of hybrid MD-*k*NN method for various depth functions. The value of $\beta = \frac{k}{N} = 0.005$.

2) Comparison of Localization Accuracy and Probability of Resolution: Localization accuracy and POR are shown for all the methods using different depth functions in Table I. Maximum depth method has the highest localization inaccuracy. *k*NN method performs well, because it accounts for the outliers. Symmetrization involved in defining depth-based neighbourhood makes *k*NN method robust to outliers as shown in Table I. The localization error and POR for hybrid MD-*k*NN is reasonable as shown in Table II. *k*NN method has highest computational complexity as it involves symmetrisation and depth-based neighbourhood in addition to what involved in maximum depth method shown in Table II.

3) Variation of Localization Error with Threshold: In order to assess the performance of the hybrid MD-*k*NN method, Figure 4 shows the variation of localization error with varying threshold, τ. As the threshold increases, localization error reduces. At larger threshold, kNN method dominates over maximum depth method. Mahalanobis, Oja and L_1 depth functions are based on absolute distances between points. Though distance based methods yields easy procedures for calculation of depth but fails to take cognizance of geometric features, which are often inherent in multivariate data. However, location and simplicial depth functions are based on relative positioning of points, it takes into account the geometric features more effectively. Hence, localization error computed using location and simplicial depth functions is minimal as shown in Figure 4. Oja depth remains constant until a particular threshold, τ = 0.96, as is expected.

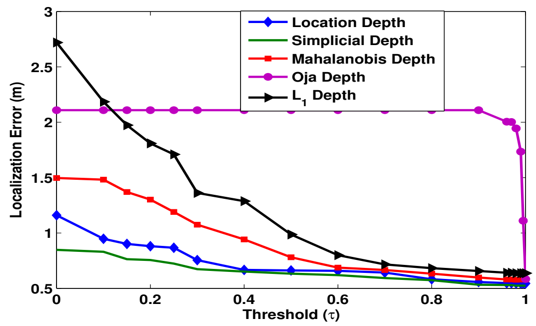

Fig. 4. Figure illustrating the variation of localization error (m) with threshold of hybrid MD-kNN method for various depth functions. The value of $\beta = 0.005$

It can be noted when performance of the hybrid MD-kNN method using various depth functions are compared in terms of computation time and localization accuracy. It is found that localization accuracy of proposed method using simplicial and location depth function is better and the requirement of computation time is longer and average respectively. On the contrary, computation time of the proposed method using Mahalanobis and L_1 depth functions is smaller, while their localization accuracy are good and average respectively.

4) Variation of Probability of Resolution with Threshold and Time: Figure 5 illustrates the variation of POR for the hybrid MD-kNN method using different depth functions. As the threshold and time increases, kNN method becomes more prominent and thus, increases the POR. Online observation vector of node improves its location estimate over time. POR using the simplicial and location depth functions outperforms other methods.

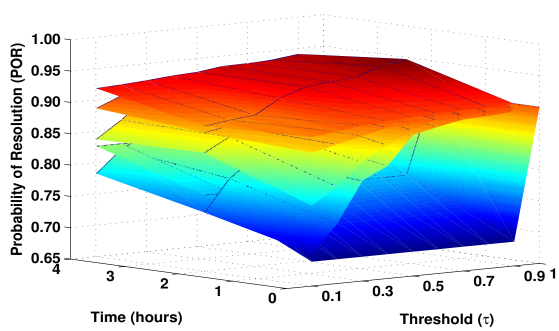

Fig. 5. Figure illustrating the variation of probability of resolution with threshold and time of hybrid MD-kNN method, for $\beta = \frac{k}{N} = 0.005$. The depth functions used are simplicial, location, Mahalanobis, Oja and L_1 from top to bottom respectively on time axis.

V. Conclusion and Future Work

In this paper, a hybrid MD-kNN method for real time sensor node tracking has been proposed. It is found that this method is inexpensive and more accurate for sensor node localization over Intel Berkeley research database. Future works includes developing an efficient tracking algorithm using long-term temporal information over AHSN. The method based on such functional data can be constructed using band depth. Few sources may provide malicious information over a time period during localization. The algorithm which accurately predicts the location of node with malicious sources is also currently being investigated. Additionally, the detection of malicious nodes can also be explored to provide a robust framework for the node tracking applications over AHSN.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications magazine, IEEE*, vol. 40, no. 8, pp. 102–114, 2002.

[2] E. D. Manley, H. Al Nahas, and J. S. Deogun, "Localization and tracking in sensor systems." in *SUTC (2)*, 2006, pp. 237–242.

[3] T. He, P. Vicaire, T. Yan, L. Luo, L. Gu, G. Zhou, R. Stoleru, Q. Cao, J. A. Stankovic, and T. Abdelzaher, "Achieving real-time target tracking usingwireless sensor networks," in *Real-Time and Embedded Technology and Applications Symposium, 2006. Proceedings of the 12th IEEE*. IEEE, 2006, pp. 37–48.

[4] P. Brida, J. Duha, and M. Krasnovsky, "On the accuracy of weighted proximity based localization in wireless sensor networks," in *Personal Wireless Communications*. Springer, 2007, pp. 423–432.

[5] V. Moghtadaiee and A. Dempster, "Indoor location fingerprinting using fm radio signals," *Broadcasting, IEEE Transactions on*, vol. 60, no. 2, pp. 336–346, June 2014.

[6] G. Ding, Z. Tan, J. Zhang, and L. Zhang, "Regional propagation model based fingerprinting localization in indoor environments," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on*. IEEE, 2013, pp. 291–295.

[7] R. Y. Liu, J. M. Parelius, K. Singh *et al.*, "Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh)," *The annals of statistics*, vol. 27, no. 3, pp. 783–858, 1999.

[8] O. Vencálek, "Concept of data depth and its applications," *Acta Universitatis Palackianae Olomucensis. Facultas Rerum Naturalium. Mathematica*, vol. 50, no. 2, pp. 111–119, 2011.

[9] Y. Zuo and R. Serfling, "General notions of statistical depth function," *Annals of statistics*, pp. 461–482, 2000.

[10] S. Kumar, S. Tiwari, and R. Hegde, "3-d mobile node localization using constrained volume optimization over ad-hoc sensor networks," in *Communications, Twentieth National Conference on*, 2014, pp. 1–6.

[11] S. Dutta and A. K. Ghosh, "On robust classification using projection depth," *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 3, pp. 657–676, 2012.

[12] D. Paindaveine and G. Van Bever, "Nonparametrically consistent depth-based classifiers," *arXiv preprint arXiv:1204.2996*, 2012.

[13] J. Hugg, E. Rafalin, K. Seyboth, and D. L. Souvaine, "An experimental study of old and new depth measures." in *ALENEX*. SIAM, 2006, pp. 51–64.

[14] J. L. Hodges, "A bivariate sign test," *The Annals of Mathematical Statistics*, pp. 523–527, 1955.

[15] J. W. Tukey, "Mathematics and the picturing of data," in *Proceedings of the international congress of mathematicians*, vol. 2, 1975, pp. 523–531.

[16] R. Y. Liu *et al.*, "On a notion of data depth based on random simplices," *The Annals of Statistics*, vol. 18, no. 1, pp. 405–414, 1990.

[17] M. A. Djauhari and R. F. Umbara, "A redefinition of mahalanobis depth function," *Malaysian Journal of Fundamental and Applied Sciences*, vol. 3, no. 1, 2008.

[18] Y. Vardi and C.-H. Zhang, "The multivariate l1-median and associated data depth," *Proceedings of the National Academy of Sciences*, vol. 97, no. 4, pp. 1423–1426, 2000.

[19] "Intel berkeley research lab data." [Online]. Available: http://db.csail.mit.edu/labdata/labdata.html

[20] N. H. Mustafa, H. R. Tiwary, and D. Werner, "A proof of the oja depth conjecture in the plane," *Computational Geometry*, vol. 47, no. 6, pp. 668–674, 2014.

[21] J. Hugg, E. Rafalin, and D. Souvaine, "Depth explorera software tool for analysis of depth measures," in *15th Annual Fall Workshop on Computational Geometry and Visualization*. Citeseer, 2005, p. 9.