ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

Device Position Localization Using Audio Fingerprint Based on Multi Signal Clustering

P.Sandhya Rani [1], P.Mohanamma [2]

M.Tech Student, Dept. of Electronics & Communication Engineering, Sree Rama Engineering College, Tirupati, India[1]

Assistant Professor, Dept of Electronics & Communication Engineering, Sree Rama Engineering College, Tirupati, India [2]

ABSTRACT: To estimate the position of device, a novel algorithm is proposed using audio fingerprinting criteria. Here by taking audio signals as sample data and using hierarchal tree cluster, audio signals are divided into clusters. Then extract the features of clusters using short time energy distribution and find similarities between clusters and database signals, based on matched points devices are localized. The proposed hierarchical tree clustering method is compared with state of art criteria like TDOA scheme and prove thet our method yields better results.

KEYWORDS: Audio fingerprinting, multi-source, hierarchical tree clustering(HTC),self-localization, time difference of arrival (TDOA) estimation.

I.INTRODUCTION

Ad-hoc sensor networks composed of randomly distributed and independent recording devices, such as smart phones, wireless microphones, and laptops, have been attracting increased interest due to their flexibility in sensor placement [1], [2]. However, the geometrical configuration of an ad-hoc array is generally unknown and may change with time. Device localization is a very important task in this context. Although not necessary in blind source separation [3] and adaptive beam forming [4], a precise knowledge of the device locations is still required in many applications such as fixed beam forming, source localization and tracking [5]. A straightforward approach for device localization is to measure the distance among all the device pairs and to apply closed-form estimators to recover their spatial locations [6].The inter-device distance can be directly measured using tapes or laser pointers, which is time consuming [10]. The inter-device distance can also be computed based on the acoustical transfer delays between two devices, by actively communicating calibration sounds between independent devices .Active communication requires a specially designed network interface or software, which is not available in many recording devices. Passive device localization instead exploits external acoustic events emitted from discrete spatial positions. Two types of information, time of arrival (TOA) and time difference of arrival (TDOA), can be estimated from microphone recordings and be used to jointly localize sensors and sources. If the source signal is known beforehand (predefined sound), its TOA for each individual microphone is obtained through cross-correlation with the given signal. If the source signal is unknown (ambient sound), the TDOA for a pair of microphones is estimated through cross-correlation of the two microphone signals. A challenge that arises for an ad-hoc array is that its devices are usually not synchronized. As a result, practical TOA measurements are biased because they may include an unknown source onset time and an unknown device capture time. Similarly, practical TDOA measurements may include an unknown time offset between a pair of recordings

Self-localization for wireless sensor networks in general is a well-studied topic. In such networks, sensor nodes communicate with neighboring nodes in a peer-to-peer manner to exchange TOA, received signal strength or angle of arrival measurements. The nodes are typically equipped with specialized RF sensing hardware for the respective modalities. Ad hoc arrays are always asynchronously sampled and hence self-localization systems based on time measurements must jointly estimate the respective local time frames w.r.t. a reference channel.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

II. EXISTING METHOD

A. Audio Landmark and Single-Source TDOA

Landmark-based audio fingerprinting is generally used for coarsely synchronizing audio recordings. However, the extracted landmark features contain some valuable information about the TDOA information of the sound sources. Without loss of generality, we consider two microphones I and j. The classical audio landmark fingerprinting converts a time domain signal $a_i(t)$ intoa sparse high-dimensional discrete time landmark feature set $F_i(n)$ At first, the time-domain signal $a_i(t)$ is transformed using the short-time Fourier transform (STFT) $A_i(n,k)$, where is the frequency index and is the frame index, which downsamples the time axis via the STFT hop size . Next, local spectral peaks $B_i(n_p,k_q)$ are selected from the power spectral amplitude $|A_i(n,k)|^2$ by comparing it with a threshold surface $\gamma(n,k)$, where n_p and k_q denotethe frame and frequency indices of the detected local peak,respectively. The threshold is initialized by the peaks found inthe first few frames. Then at each frame is updated by a decaying factor and is also raised by peaks found in the previous frame. All the local peak values higher than the threshold are kept in with other peaks set to zeros ('pruning').The threshold is updated as

$$\gamma(n,k)=max\left(\frac{\gamma(n-1,k)}{\beta}, G(k) * B_i(n,.)\right)$$

Where G(k) is a Gaussian function and denotes the convolution operator in frequency. The number of local peaks is controlled by $\beta$and the variance,σ , of G(k) . The larger β and σ the more local peaks will be selected.
A landmark, y $(n_1,k_1;n_2,k_2)$, is formed by pairing up twonearby local spectral peaks $B_i(n_1,k_1)$and $Bi(n_2,k_2)$. To reducethe dimension, each landmark is hashed into an integervalue using

$$F = k_1 2^{12} + (k_2 - k_1)2^6 + (n_2 - n_1)$$

In this way, the obtained landmarks associated with the time frame are represented as a time-indexed feature set $F_i(n) = \{F_u\}_{u=1}^{U_n}$,where U_n is the total number of landmarks at the frame .The extracted audio landmarks contain the TDOA information of the sound sources. Assume only the b-th source is active and the time offset between two microphones is zero. Then the STFT of $a_i(t)$and $a_j(t)$can be expressed

$$A_i(n,k) = g_{ib}S_b(n - n_{ib,k}),$$
$$A_j(n,k) = g_{jb}S_b(n - n_{jb,k})$$

where $S_b(.)$is the STFT of ,$S_b(.), n_{ib} = \lfloor t_{ib}/R \rceil$, and$n_{jb} = \lfloor t_{jb}/R \rceil$ where$\lfloor . \rceil$ the operator denotes the integer part.By landmark matching between the two channels [12], thelandmarks corresponding to the same time-frequency peak pairscan be extracted:

$$y_i(n_1 - n_{ib}, k_1; n_2 - n_{ib}, k_2) = y_j(n_1 - n_{jb}, k_1; n_2 - n_{jb}, k_2)$$

and consequently

$$F_i(n) = F_j\left(n - \left\lfloor \frac{T_{ijb}}{R} \right\rceil\right)$$

Where$F_i(.)$ and $F_j(.)$denote the extracted audio fingerprints of $a_i(.)$and $a_j(.)$, respectively; and$y_i(.)$ and$y_j(.)$ denote two matched local peak pairs in the two channels. The time delay between two channels and the matched landmarks are clearly related, where the hop size determines the resolution of the time delay.

As example, in a simulated anechoic environment a sound source (speech) is placed at an end-fire location with respect to a pair of (synchronized) microphones which are apart. The audio fingerprint extraction procedure is visualized in Fig. 3, where the spectrogram of the speech signal, the pruning threshold surface, and the extracted audio landmarks in the first channel are depicted in the three subfigures. The audio fingerprint matching results between two channels are visualized in Fig. 4, where four examples of matched audio landmarks in a short segment (5.8s-6.0s) in the two channels are depicted. The matched landmarks in the two channels typically occur at the same frequency bins but with a temporal offset of 0.0145s,which equals to the acoustic transmission time of 5ms .

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

B. EXTREME TDOA ESTIMATION METHOD

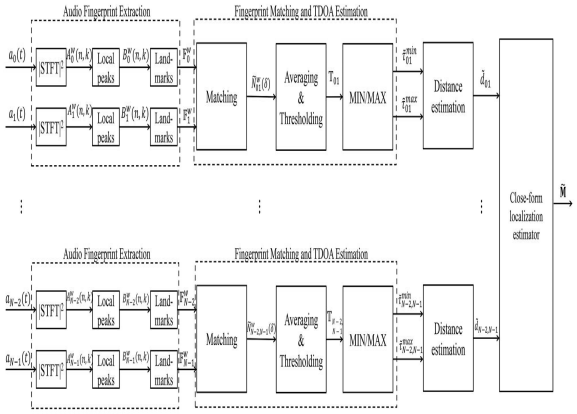

Fig:1 The block diagram of the audio-fingerprinting-based on TDOA estimation algorithm.

Based on the analysis above, we propose an audio-fingerprinting based method to estimate the extreme TDOAs from a multi-source environment and then utilize the estimated pair-wise distances to compute the device locations in fig.1. The signals, $a_i(t), i = 0, \ldots \ldots$,N-1 are divided into non overlapping segments, $a_i^w(t), w = 1, \ldots, w$ of length. The time domain signal $a_i^w(t)$ is transformed to the audio landmark feature set $F_i^w(n)$. In each segment we calculate the number ofmatched audio landmarks of $F_i^w(n)$ and $F_j^w(n)$ at different timeshifts δ:

$$N_{ij}^w(\delta) = \sum_{n=1}^{\lfloor L/R \rfloor - 1} \langle F_i^w(n) \cap F_j^w(n+\delta) \rangle$$

Where $\cap$ is the intersection and $\langle . \rangle$ is the cardinality of a set. The number of matched landmarks $\bar{N}_{ij}^w(\delta)$ is averaged across allthe segments, obtaining

$$\bar{N}_{ij}(\delta) = \frac{1}{W} \sum_{w=1}^{W} \bar{N}_{ij}^w(\delta)$$

A small number of outliers randomly distributed across different segments may still exist in $\bar{N}_{ij}^w(\delta)$ due to some mismatched landmarks. Averaging $\bar{N}_{ij}^w(\delta)$ over all segments helps to suppressthese outliers. We refer to $\bar{N}_{ij}(\delta)$ as the matching score between two channels. When only one source exists, e.g. the b-th source, its TDOA can be estimated in a similar way as GCC, i.e.,

$$T_{ij} = \text{argmax}\{\tilde{N}_{ij}(\delta)\}.R$$

Where δ is in the range $\{-\lfloor \frac{L}{R} \rfloor, \ldots \ldots, \lfloor \frac{L}{R} \rfloor\}$ with a step of 1.When multiple sources exist, $N_{ij}(\delta)$ is seen as a histogram ofthe TDOAs of these sources. We remove the residual outliers by applying a threshold α to $N_{ij}(\delta)$ and estimatea set of TDOAs T_{ij} by using

$$T_{ij} = \{R.\delta | \bar{N}_{ij}(\delta) > \alpha\}$$

The minimum TDOA $\bar{T}_{ij}^{min}$ and maximum TDOA $\bar{T}_{ij}^{max}$ can beestimated as

$\bar{T}_{ij}^{min}$=min (T_{ij}) and $\bar{T}_{ij}^{max}$=max (T_{ij})

The inter-device distance $\bar{d}_{ij}$ is calculated using $\bar{T}_{ij}^{min}$ and $\bar{T}_{ij}^{max}$, as given. In the same way the distance of each pair of devices in the ad-hoc array can be estimated. The pair-wise distances are further used to recover the geometrical configuration of the array, based on the closed-form estimator.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

III. PROPOSED MITIGATION SCHEME

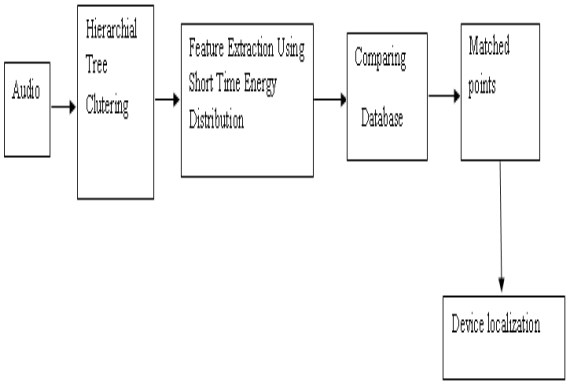

Fig:2. The block diagram of the proposed method

The proposed block diagram in fig.2 describes the overall process. First the test audio is taken then the Hierarchal Tree Clustering is applied to analyze the data. The feature is extracted from the data by using short time energy distribution. Then by comparing with the database we get the matched points. By the matched points we find the location of the device.

I. Hierarchal Tree Clustering

Hierarchical clustering is a widely used data analysis tool. The idea is to build a binary tree of the data that successively merges similar groups of points. Visualizing this tree provides a useful summary of the data. Each level of the resulting tree is a segmentation of the data. The algorithm results in a sequence of groupings. It is up to the user to choose a ”natural” clustering from this sequence. Agglomerative clustering is monotonic.

The similarity between merged clusters is monotone decreasing with the level of the merge. The agglomerative hierarchical clustering algorithms available in this program module build a cluster hierarchy that is commonly displayed as a tree diagram called a dendrogram. They begin with each object in a separate cluster. At each step, the two clusters that are most similar are joined into a single new cluster. Once fused, objects are never separated. The horizontal axis of the dendrogram represents the distance or dissimilarity between clusters. The vertical axis represents the objects and clusters. The dendrogram is fairly simple to interpret. Remember that our main interest is in similarity and clustering. Each joining (fusion) of two clusters is represented on the graph by the splitting of a horizontal line into two horizontal lines. The horizontal position of the split, shown by the short vertical bar, gives the distance (dissimilarity) between the two clusters.

The algorithm used by all eight of the clustering methods is outlined as follows. Let the distance between clusters i and j be represented as d_{ij} and let cluster i contain n_i objects. Let D represent the set of all remaining d_{ij} . Suppose there are N objects to cluster. 1. Find the smallest element d_{ij} remaining in D. 2. Merge clusters i and j into a single new cluster, k. 3. Calculate a new set of distances d_{km} using the following distance formula.

$$d_{km} = \alpha_i d_{im} + \alpha_j d_{jm} + \beta d_{ij} + \gamma \left| d_{im} - d_{jm} \right|$$

Here m represents any cluster other than k. These new distances replace dim and d_{jm} in D. Also let $n_k = n_i + n_j$. Note that the eight algorithms available represent eight choices for α_i ,α_j , β, and γ . 4. Repeat steps 1 - 3 until D contains a single group made up off all objects. This will require N-1 iterations. We will now give brief comments about each of the eight techniques.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

IV.RESULTS

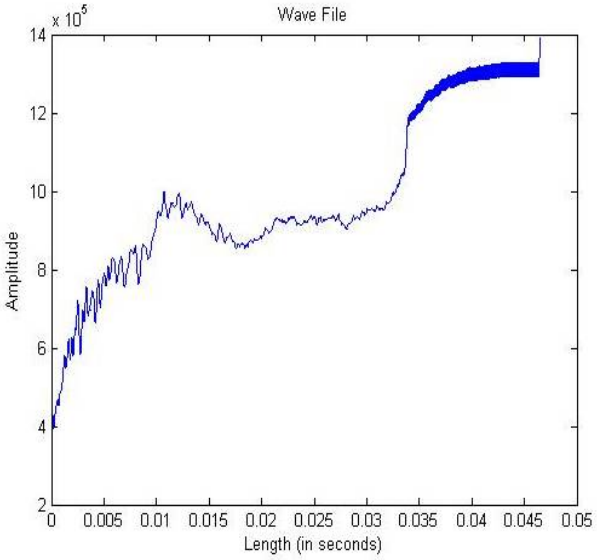

FIG.1: Audio signal 1

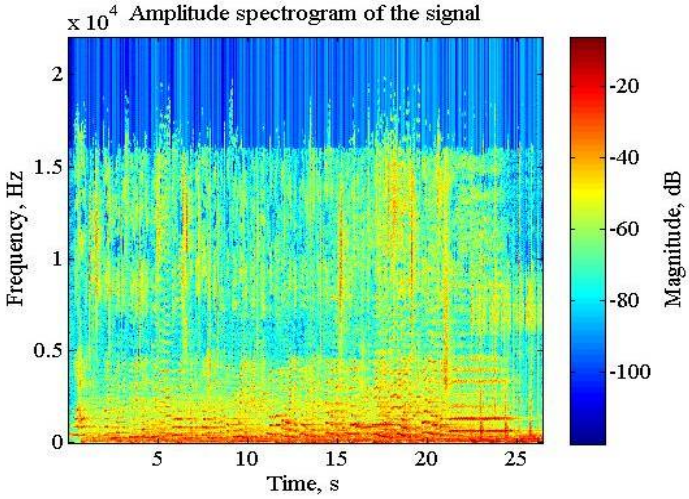

FIG.2: Amplitude Spectrum of the Signal 1

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

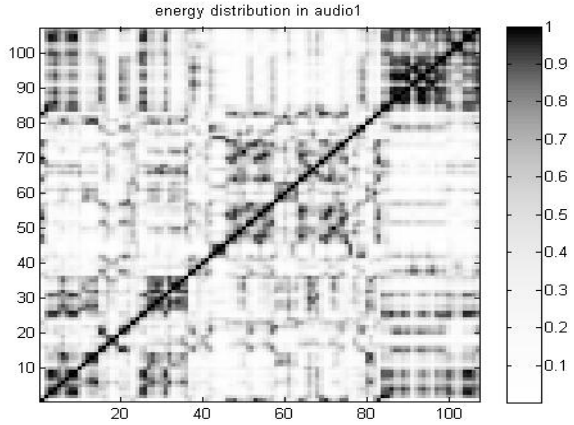

FIG.3: Energy Distribution in Audio1

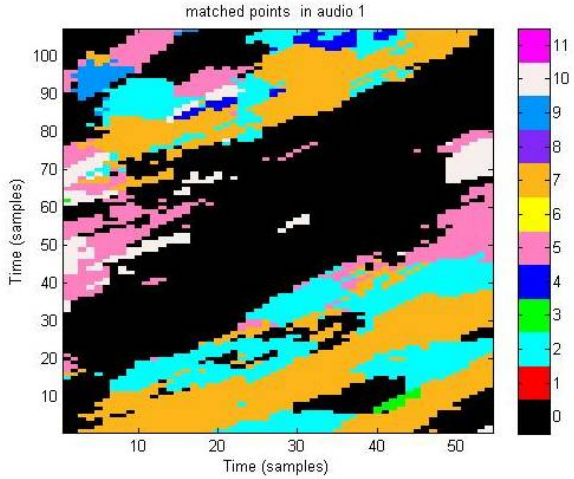

FIG.4: Matched Points in Audio1

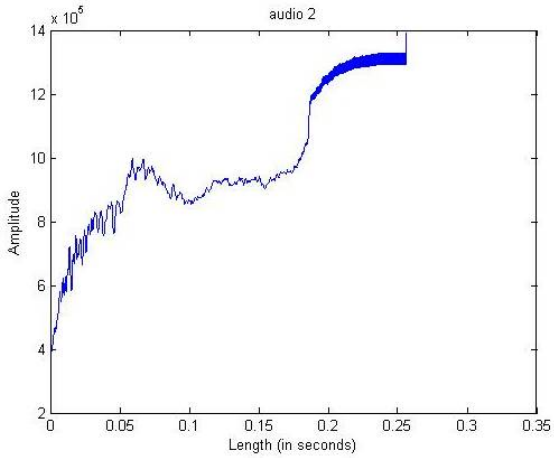

FIG.4: Audio Signal 2

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

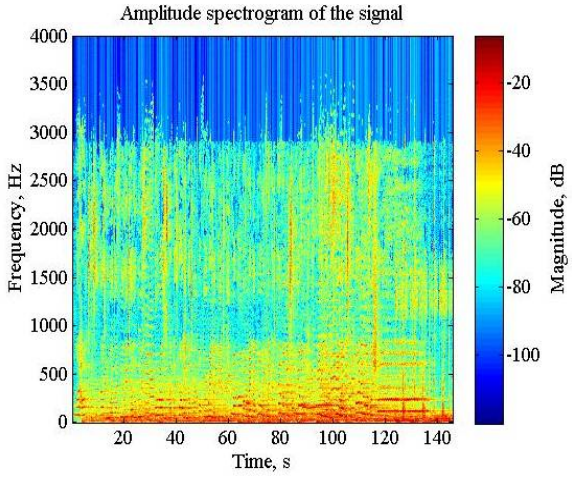

FIG.4:Amplitude spectrum of the Audio signal 2

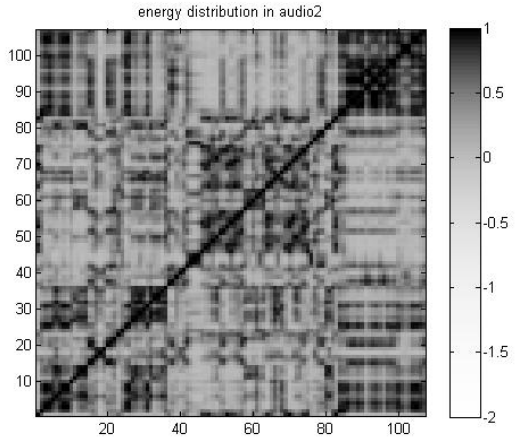

FIG.5: Energy Distribution in Audio2

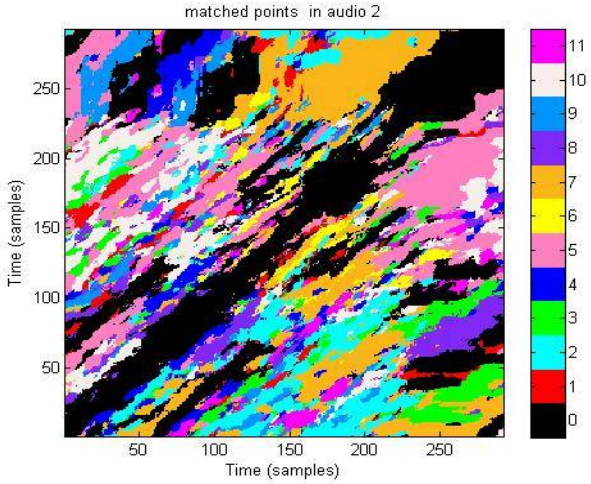

FIG.6: Matched Points in Audio2

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

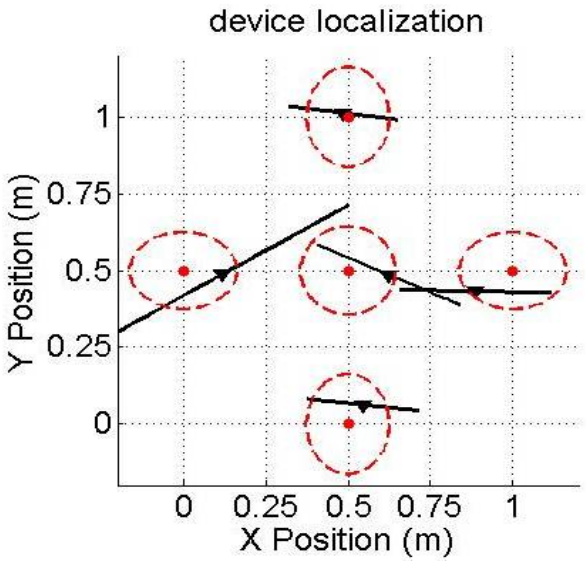

FIG.7: Device Localization

V.CONCLUSION

Device self-localization problem is addressed in an ad-hoc sensor network by exploiting the clusters from multiple sound sources to asynchronous devices. In the proposed method for device localization using audio fingerprinting criteria, here the audio signals are taken as sample data and using hierarchal tree cluster then divided the audio signal into clusters. Then extracted the features of clusters using short time energy distribution function and find similarities between clusters and database signals, based on matched points we localized the devices. Using extracted landmark features consisting of pairs of spectral peaks of the audio signal for hierarchical tree clustering was found to be more robust to noise than the TDOA estimation algorithm. We compared our method with state of art criteria like TDOA scheme and proved that proposed method yields better results.

REFERENCES

[1] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell,"A survey of mobile phone sensing," *IEEE Commun. Mag.*, vol.48, no. 9, pp. 140–150, Sep. 2010.
[2] M. Hennecke and G. Fink, "Towards acoustic self-localization ofad hoc smartphone arrays," in *Proc. Joint Workshop Hands-FreeSpeech Commun. Microphone Arrays*, Edinburgh, U.K., May 2011,pp. 127–132.
[3] H. Aghajan and A. Cavallaro, *Multi-camera networks: Principles andapplication*. New York, NY, USA: Academic, 2009.
[4] N. Drawil, H. Amar, and O. Basir, "GPS localization accuracy classification:A context-based approach," *IEEE Trans. Intell. Transportat.Syst.*, vol. 14, no. 1, pp. 262–273, Mar. 2013.
[5] N. Ravi, P. Shankar, A. Frankel, A. Elgammal, and L. Iftode, "Indoorlocalization using camera phones," in *Proc. IEEE Workshop MobileComput. Syst. Appl.*, Orcas Island, Aug. 2006, vol. Suppl., pp. 1–7.
[6] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "Beepbeep: A highaccuracy acoustic ranging system using cots mobile devices," in *Proc.ACM Int. Conf. Embedded Netw. Sensor Syst.*, Sydney, Australia, Nov.2007, pp. 1–14.
[7] J. Schmalenstroeer, P. Jebramcik, and R. Haeb-Umbach, "A combinedhardware-software approach for acoustic sensor network synchronization,"*Signal Process.*, vol. 107, pp. 171–184, Feb. 2015.
[8] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannelsampling frequency mismatch for ad hoc microphone arraybased on maximum likelihood estimation," *Signal Process.*, vol. 107,pp. 185–196, Feb. 2015.
[9] J. Bahi, A. Makhoul, and A. Mostefaoui, "A mobile beacon based approachfor sensor network localization," in *Proc. IEEE Int. Conf.Wireless Mobile Comput., Netw., Commun.*, White Plains, NY, USA, Oct.2007, p. 44.
[10] P. Pertila, M. Mieskolainen, and M. Hamalainen, "Passive self-localizationof microphones using ambient sounds," in *Proc. Eur. SignalProcess. Conf.*, Bucharest, Romania, Aug. 2012, pp. 1314–1318.
[11] P. Pertila, M. Hamalainen, and M. Mieskolainen, "Passive temporaloffset estimation of multichannel recordings of an ad-hoc microphonearray," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 11, pp.2393–2402, Nov. 2013.
[12] A. L. Wang, "An industrial-strength audio search algorithm," in *Proc.Int. Conf. Music Info. Retrieval*, Baltimore, MD, USA, Oct. 2003, pp.

ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 6, Special Issue 2, February 2017

National Conference on Recent Advancements in Engineering & Technology and Management 2017 (NCRAETM 2K17)

18th February 2017

Organized by

Sree Rama Engineering College (SRET), Tirupati- 517 507, India

7–13.
[13] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures viatime-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7,pp. 1830–1847, Jul. 2004.HON
[14] K. Chintalapudi, A. Padmanabha Iyer, and V. N. Padmanabhan, "Indoorlocalization without the pain," in *Proc. ACM Int. Conf. MobileComput. Netw.*, Chicago, IL, USA, Sep. 2010, pp. 173–184.
[15] A. Tsui, W.-C. Lin, W.-J. Chen, P. Huang, and H.-H. Chu, "Accuracyperformance analysis between war driving and war walking inmetropolitan Wi-Fi localization," *IEEE Trans. Mobile Comput.*, vol. 9,
no. 11, pp. 1551–1562, Nov. 2010.
[16] Z. Zhong and T. He, "MSP: Multi-sequence positioning of wirelesssensor nodes," in *Proc. ACM Int. Conf. Embed. Netw. Sens. Syst.*,Sydney, Australia, Nov. 2007, pp. 15–28.
[17] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher,"Range-free localization schemes for large scale sensor networks," in*Proc. ACM Int. Conf. Mobile Comput. Netw.*, San Diego, CA, USA,
Sep. 2003, pp. 81–95.
[18] D. Niculescu and B. Nath, "DV based positioning in ad hoc networks,"*Tele. Syst.*, vol. 22, pp. 267–280, Jan. 2003.
[19] P. Robertson, M. Angermann, and B. Krach, "Simultaneous localizationand mapping for pedestrians using only foot-mounted inertial sensors,"in *Proc. ACMInt. Conf.Ubiquitous Comput.*, Orlando, FL, USA,Sep. 2009, pp. 93–96.
[20] O. Woodman and R. Harle, "Pedestrian localization for indoor environments,"in *Proc. ACM Int. Conf. Ubiquitous Comput.*, Seoul, Korea,
Sep. 2008, pp. 114–123.
[21] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motioncausally integrated over time," *IEEE Trans. Pattern Anal.Mach. Intell.*,
vol. 24, no. 4, pp. 523–535, Apr. 2002.
[22] B. Williams, G. Klein, and I. Reid, "Real-time SLAM relocalization,"in *Proc. IEEE Int. Conf. Comput. Vis.*, Rio, Brazil, Oct. 2007, pp. 1–8.
[23] A. Davison, I. Reid, N. Molton, and O. Stasse, "MonoSLAM: Realtimesingle camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*,vol. 29, no. 6, pp. 1052–1067, Jun. 2007.
[24] M. Hebert, "Active and passive range sensing for robotics," in *Proc.IEEE Int. Conf. Roboti. Autom.*, San Francisco, CA, USA, 2000, vol.1, pp. 102–110.
[25] N. Anjum and A. Cavallaro, "Automated localization of a camera network,"*IEEE Intell. Syst.*, vol. 27, no. 5, pp. 10–18, Sep. 2012.