SOURCE LOCALIZATION FROM QUANTIZED TIME OF ARRIVAL MEASUREMENTS

Nadir Castañeda, Maurice Charbit and Eric Moulines

CNRS LTCI & GET / Télécom Paris,
46 rue Barrault, 75634 Paris cedex 13, France

ABSTRACT

In this paper, we consider the localization of a source from quantized measurements of time of arrivals (TOA) or time difference of arrivals (TDOA). Applications include, as particular examples, acoustic source localization from a network of microphones under communication constraints, and the localization of a base station using a geolocalized mobile station using timing advance measurements. We use a maximum likelihood approach, based on an efficient implementation of the EM algorithm. Contrary to previously reported work, our technique takes into account not only the measurement noise, but also the presence of outliers (for example, non line of sight propagation) and the quantization. We illustrate our findings using simulated data and real field measurements.

1. INTRODUCTION

This communication is concerned with the problem of source localization using a network of geolocalized sensors. We assume that each sensor of this network measures signals generated from a single source assumed to stay at a given position in space during all the localization procedure. The nature of the measurements depends on the modalities which are used but the statistical model we propose is flexible enough to cover most situations of interest. In the sequel we focus on two scenarios, the first one based on time of arrival (TOA) and the second one on time difference of arrivals (TDOA); other types of measurements (e.g. direction of arrival) fit in the same framework.

Consider first the localization of an acoustic source with M synchronized microphones. In this type of experiments, the recorded signal on each microphone is decomposed into frames (approx. 20 msec). For each pair (i, j) of microphone and each frame, we evaluate the maximum of the normalized cross-correlation, which gives a measure of the time difference of arrival (TDOA). In absence of noise and quantization effects, the TDOA is given by $\rho_{i,j}(x_s) = c^{-1}(|x_s - x_i| - |x_s - x_j|)$ where x_s denotes the unknown location of the source, x_i the (known) location of the i-th microphone and c the speed of sound.

The TDOA measurement is typically affected by several sources of impairment. First, because the input signal of each microphone is recorded in presence of noise, the maximum of the cross-correlation is always affected by a random fluctuation, even if the direct path is dominant. This fluctuation depends on the SNR of the microphone pair. For simplicity, as suggested in [3], this error may be modelled as an additive Gaussian noise. Second, if the SNR on a given microphone falls below a threshold or in the presence of a secondary path, the maximum of the correlation will be significantly away from the true TDOA and the observation will contain almost no useful information on the source localization; these values must be interpreted as an outlying observation. As studied in [3] in a typical acoustic source localization problem, the number of outliers depends heavily on the SNR ranging from 5% for high SNR to 35% for low SNR (section 2.1 in [3]). For the outlying observations an additive noise model is not appropriate. The approach we propose is to model the outliers as a gaussian variable typically with high variance. Finally, the cross-correlation is evaluated on a grid which introduces a quantization error. If the communication constraints are stringent (implying transmission of significantly down-sampled version of the input signal) or if the computational complexity is limited, the quantization error might be significant and should be taken into account.

To sum up, the observation for each pair of microphones is given by

$$\begin{cases} Z_{i,j} \sim \gamma_1 \mathcal{N}(\rho_{i,j}(x_s), \sigma_1^2) + \gamma_2 \mathcal{N}(0, \sigma_2^2) \\ Y_{i,j} = Q(Z_{i,j}) \end{cases} \tag{1}$$

where Q is the quantizer, σ_1^2 the variance of the measurement noise for "normal" (not outlying) observation, σ_2^2 the variance of the outlying observation, and γ_2 the proportion of outliers. $\mathcal{N}(\mu, \sigma^2)$ denotes the gaussian distribution with mean μ and variance σ^2.

A model similar to (1) can be used for localization using TOA. An interesting example is given by the localization of a fixed or a mobile station in a TDMA based wireless network. In TDMA based systems, such as GSM, the base station (BS) send, for synchronization purposes, to each mobile station (MS) a timing advance (TA) which represents the perceived amount of round-trip propagation delay BS-MS-BS. In this context, localization may be performed using this TA information. Once again, the round-trip time is obtained by computing the maximum of the correlation on a burst sent by the BS and echoed by the MS. In presence of direct path, the measured round-trip delay is directly related to the distance between the BS and the MS. This measure is subject to random error which can be modelled by an additive Gaussian noise. Notice that, because unknown timing offset in the MS, the noise is not necessary zero-mean. On the other hand, in presence of non line of sight propagation (NLOS), the measurement does not carry any useful information on the distance MS-BS. Therefore, measurements obtained in presence of NLOS propagation paths can be considered as outliers. Finally, in the GSM standard, the TA is quantized to save bandwidth. The quantization function is defined as

$$Q(z) = \begin{cases} 0 & \text{if} \quad 0 \leq z < \frac{q}{2} \\ y & \text{if} \quad (y - \frac{1}{2})q \leq z < (y + \frac{1}{2})q \end{cases} \tag{2}$$

where $y \in \{1, \cdots, 63\}$ and where $q = 553.8$m denotes the quantization step. We are back to a model similar to (1), this time the observed TOA model is $Z_i \sim \gamma_1 \mathcal{N}(\mu_1 + d_i(x_s), \sigma_1^2) + \gamma_2 \mathcal{N}(\mu_2, \sigma_2^2)$ where $d_i(x_s)$ is the distance between the source in the location x_s and the i-th sensor in the known location x_i and where μ is some offset value.

We believe that similar models featuring errors, outliers and quantization will be useful in many others situations (details will be given in an extended version of this proceeding paper).

2. ALGORITHM DERIVATION

We consider a sequence $\{Y_1, \cdots, Y_n\}$ of n observations taking its values in the finite set $\{1, \cdots, J\}$. Referring to (1), we assume that these observations are the quantized versions of a sequence $\{Z_1, \cdots, Z_n\}$ of n independent random variables distributed as a Gaussian mixture model. Because direct maximization of the likelihood of the observations is intractable, we suggest to use EM approach. The EM algorithm [1] is a very popular tool for maximum-likelihood (or maximum a posteriori) estimation. The common strand to problems where this approach is applicable is a notion of *incomplete-data*, which includes the conventional sense of missing data but is much broader than that. The EM algorithm demonstrates its strength in situations where some hypothetical experiments yield (complete) data that are related to the parameters more conveniently than the measurements are.

According to the model introduced above, the joint probability density of the complete data is given by

$$\begin{aligned} f(z_i, u_i; \theta) \quad &= \quad \gamma_1 \phi(z_i; \chi_i(x_s) + \mu_1, \sigma_1^2) \mathbb{1}(u_i = 1) \\ &\quad + \gamma_2 \phi(z_i; \mu_2, \sigma_2^2) \mathbb{1}(u_i = 2) \end{aligned} \tag{3}$$

where $\chi_i(x_s)$ is a known function depending upon the location x_s of the source to be located and the location of the i-th sensor (or pair of sensors for TDOA), μ_1 a constant offset, $0 < \gamma_2 < 1$ the proportion of outliers ($\gamma_1 = 1 - \gamma_2$) and $\phi(z; \mu, \sigma^2)$ the pdf of a Gaussian with mean μ and variance σ^2. The full parameter vector is denoted $\theta = \{x_s, \gamma_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2\}$.

The EM algorithm is an iterative algorithm to compute maximum likelihood estimate. Each iteration may be formally decomposed in two steps: an E-step and a M-step. The E-step consists in evaluating the conditional expectation of the complete data likelihood

$$Q(\theta, \theta_p) = \sum_{i=1}^{n} \mathbb{E}\{\log(p(z_i, u_i; \theta)) | y_i, \theta_p\} \tag{4}$$

where θ_p is the current fit of the parameters at p-th iteration and where the expectation is taken w.r.t. the probability distribution associated with the value θ_p of the parameter. In the (generalized) M-step, we compute a new parameter estimate, θ_{p+1}, which is chosen in such a way that $Q(\theta_{p+1}, \theta_p) \geq Q(\theta_p, \theta_p)$ with equality if and only if θ_p is a stationary point of the likelihood function. This two step process is repeated until convergence is apparent. The essence of the EM is that increasing $Q(\theta, \theta_p)$ forces an increase of the incomplete data likelihood. Let us denote

$$\begin{aligned} F(a, b) \quad &= \quad \int_a^b (2\pi)^{-1/2} e^{-v^2/2} dv, \\ G(a, b) \quad &= \quad \int_a^b (2\pi)^{-1/2} v e^{-v^2/2} dv, \\ H(a, b) \quad &= \quad \int_a^b v^2 (2\pi)^{-1/2} e^{-v^2/2} dv. \end{aligned}$$

For $1 \leq i \leq n$ and $1 \leq j \leq 2$, we let $\bar{\mathbb{P}}_i = \sum_{j=1}^{2} \tilde{\gamma}_j F(\tilde{\alpha}_{i,j}, \tilde{\beta}_{i,j})$, $\bar{F}_{i,j} = \tilde{\gamma}_j F(\tilde{\alpha}_{i,j}, \tilde{\beta}_{i,j}) / \bar{\mathbb{P}}_i$, $\bar{G}_{i,j} = \tilde{\gamma}_j G(\tilde{\alpha}_{i,j}, \tilde{\beta}_{i,j}) / \bar{\mathbb{P}}_i$ and $\bar{H}_{i,j} = \tilde{\gamma}_j H(\tilde{\alpha}_{i,j}, \tilde{\beta}_{i,j}) / \bar{\mathbb{P}}_i$ where, assuming a constant quantization step q, we have set

$$\begin{aligned} &\tilde{\alpha}_{i,1} = \tilde{\sigma}_1^{-1} (\max\{(y_i - 1/2)q, 0\} - \chi_i(\tilde{x}_s) - \tilde{\mu}_1) \\ &\tilde{\beta}_{i,1} = \tilde{\sigma}_1^{-1} ((y_i + 1/2)q - \chi_i(\tilde{x}_s) - \tilde{\mu}_1) \\ &\tilde{\alpha}_{i,2} = \tilde{\sigma}_2^{-1} (\max\{(y_i - 1/2)q, 0\} - \tilde{\mu}_2) \\ &\tilde{\beta}_{i,2} = \tilde{\sigma}_2^{-1} ((y_i + 1/2)q - \tilde{\mu}_2) \end{aligned}$$

Plugging these notations in (4) we obtain

$$\begin{aligned} Q(\theta, \tilde{\theta}) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{2} \cdots & \qquad\qquad (5) \\ \frac{1}{\sigma_j^2} \left(\tilde{\sigma}_j^2 \bar{H}_{i,j} + 2\tilde{\sigma}_j \delta_{i,j} \bar{G}_{i,j} + \delta_{i,j}^2 \bar{F}_{i,j} \right) \cdots & \\ + \log(2\pi\sigma_j^2) \bar{F}_{i,j} - 2\log(\gamma_j) \bar{F}_{i,j} & \end{aligned}$$

where $\delta_{i,1} = \chi_i(\tilde{x}_s) - \chi_i(x_s) + (\tilde{\mu}_1 - \mu_1)$ and $\delta_{i,2} = \tilde{\mu}_2 - \mu_2$. We then apply a relaxation scheme, where we first optimize w.r.t the proportions γ_i, the means μ_i and the variances σ_i^2 keeping the source location constant and then optimize w.r.t.

the source location keeping all others parameters constant.

$$\begin{aligned}\gamma_j &= \frac{1}{n}\sum_{i=1}^{n}\bar{F}_{i,j},\\ \mu_j &= \tilde{\mu}_j + \tilde{\sigma}_j\frac{\sum_{i=1}^{n}\bar{G}_{i,j}}{\sum_{i=1}^{n}\bar{F}_{i,j}},\\ \sigma_j^2 &= \tilde{\sigma}_j^2\frac{\sum_{i=1}^{n}\bar{H}_{i,j}}{\sum_{i=1}^{n}\bar{F}_{i,j}} - (\tilde{\mu}_j - \mu_j)^2.\end{aligned}$$

The maximization of $Q(\theta, \tilde{\theta})$ w.r.t. the source location, even by keeping all others parameters constant, has not closed form expression. In order to avoid a costly 2D or 3D grid search, we apply a single gradient EM step as in [2] for x_s only.

3. EXPERIMENTAL RESULTS AND COMMENTS

In our experiments, we consider a Time of Arrival Based Location Algorithm (TABLA) in the GSM scenario. The MS is moving along some known trajectory receiving TA measurements sent by a BS to be located. To assess performance of TABLA we use both simulated data and field data measurements.

3.1. Simulated data

Data are simulated from the model given by expression (3). Both timing offset of the MS (parameter μ_1) and presence of outliers from NLOS propagation paths (parameter γ_2) are considered. More precisely we take $\mu_1 = 700$ m, approximately half the quantization step and $\sigma_1 = 100$ m. The outlying component is taken to be Gaussian with mean $\mu_2 = 6000$ m and standard deviation $\sigma_2 = 450$ m. The proportion of outliers is set to $\gamma_2 = 30\%$. The MS travels in the first case along a straight line of 4 km and in second case along a circle whose diameter is 2.5 km. In both cases, the MS moves at constant speed and obtains a new value of the TA (which is computed according to the GSM specifications) respectively every 1.00 m and 2.00 m. To get a better understanding of the difficulty of problem, we have displayed the uncertainty ellipses for the parameter x_s computed from the Cramer-Rao Bound (CRB) (details of the computations are not reported here) for different BS locations (on a two-dimensional grid, with a grid width of 500m on each coordinate). In figures 1 and 2, the MS moves on a circle and on a line, respectively. These figures shows that the precision of the BS highly depends on the localization of the BS with respect to the sensors positions (points where the MS get TA measurements).

We have checked from 250 independent Monte-Carlo experiments carried out on a subset of the BS positions taken from the grid that these 95 % confidence region have an accurate correct coverage probability. This shows that, despite the fact that the likelihood surface has several local extremas, convergence occurs almost always to the global maximum of the likelihood function (multimodality is thus not a severe problem). It is worthwhile to note that the precision of the estimator is much lower than the TA quantization interval in the two scenarios, which may appear as a surprise.

3.2. A real experiment

We assess the validity of our approach and the feasibility of the BS localization from TA using real measurement. In this experiment, we use a MS which is able to force the selection of the serving BS (this function is not usually available on commercial MS) and we record synchronously the TA and the MS location using a commercial GPS system. To illustrate our findings, we consider a specific experiment carried out in a rural area. In figure 3, we have displayed the trajectory followed by the MS (green curve) and the position of the BS (yellow dot). The rings centered the BS correspond to regions where the (theoretical) TA is constant, and give a feeling of the effect of the quantization, which is extremely important in this case. In figure 4, we have displayed the measured TA as a function of the distance (the number of observations is 1148). This figure shows that there are only few outliers, but a substantial amount of noise; there is clearly a timing offset (approximately half a symbol period). The EM converges in one iteration, despite a poor initialization (the initial point of the EM is localized at 1500m from the BS (it is not displayed in figure 3) and the final iterate is at 92m from the true BS (estimated offset 441 m; estimated outliers proportion 1%). It should be stressed that, in this scenario, the EM iterates converge fast toward the region of interest (in two steps, the iterate is less than 200 m from the true BS location).

4. REFERENCES

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *Journal of Royal Statistical Society : series B*, vol. no. 39, pp. 1–38, 1977.

[2] K. Lange, “A Gradient Algorithm Locally Equivalent to the EM Algorithm,” *Journal of Royal Statistical Society : series B*, vol. no. 2, pp. 425–437, 1995.

[3] T. Ajdler, I. Kozintsev, R. Lienhart, and M. Vetterli, “Acoustic Source Localization in Distributed Sensor Networks,” *Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA*, vol. 2, pp. 1328–1332, 2004.

[4] M. Silventoinen and T. Rantalainen, “Mobile Station Locating in GSM,” *Wireless Communications System Symposium, IEEE*, pp. 53–59, Nov 1995.

[5] H.-L. Song, “Automatic vehicle location in cellular communications systems,” *IEEE Trans. Veh. Technol.*, vol. 43, no. 4, pp. 902–908, Nov 1994.

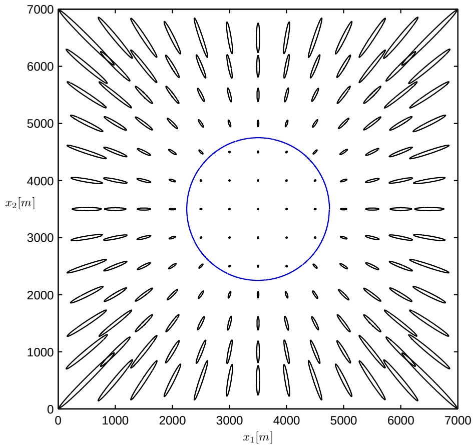

Fig. 1. Uncertainty ellipses for BS localization. Circular Motion

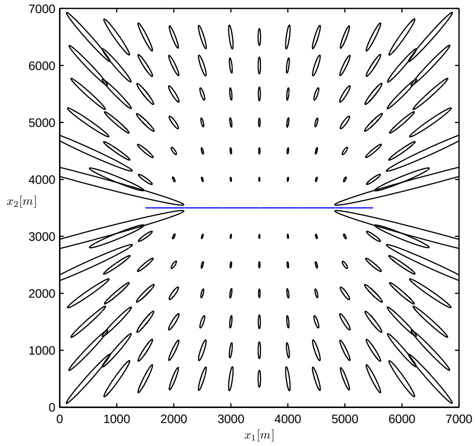

Fig. 2. Uncertainty ellipses for BS localization. Linear Motion

[6] L. Ding-Bing and J. Rong-Terng, “Mobile location estimation based on differences of signal attenuations for gsm systems,” *IEEE Trans. Veh. Technol.*, vol. 54, no. 4, pp. 1447–1454, Jul 2005.

[7] M. Spirito and A. Mattioli, “Preliminary experimental results of a gsm mobile phones positioning system based on timing advance,” *Veh. Technol. Conf.*, vol. 4, pp. 2072–2076, Sep 1999.

[8] M. A. Spirito, “On the accuracy of cellular mobile station location estimation,” *IEEE Trans. Veh. Technol.*, vol. 50, no. 3, pp. 674–685, May 2001.

[9] C. Drane, M. Macnaughtan, and C. Scott, “Positioning gsm telephones.” *IEEE Commun. Mag.*, vol. 36, pp. 46–54, Apr 1998.

[10] J. Caffery and G. Stuber, “Vehicle location and tracking for ivhs in cdma microcells,” *5th IEEE International Symposium Personal, indoor and Mobile Radio Communications*, vol. 4, pp. 1227–1231, Sep 1994.

[11] F. Cesbron and R. Arnott, “Locating gsm mobiles using antenna array.” *Electronics letters.*, vol. 34, pp. 1539–1540, Aug 1998.

[12] X. Wang, Z. Wang, and B. O’Dea, “A toa-based location algorithm reducing the errors due to non-line-of-sight (nlos) propagation.” *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, pp. 112–116, Jan 2003.

[13] ETSI., “Digital cellular telecommunications system (phase 2); radio subsystem synchronization (ets 300 579).” *ETSI, Sophia Antipolis - France*.

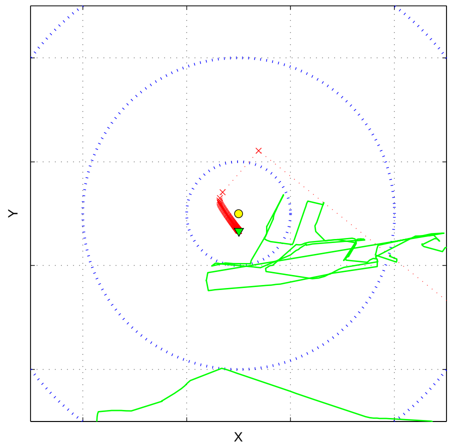

Fig. 3. Green curve: Trajectory of the MS. Yellow dot: location of the BS. Red cross: Estimates of the BS positions

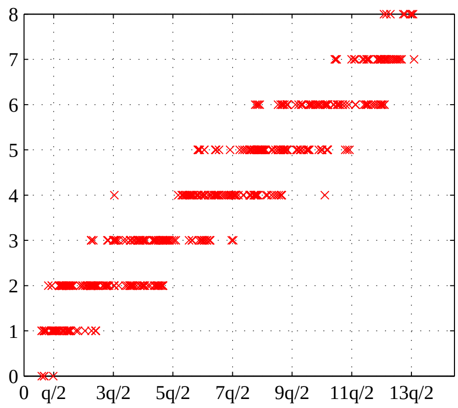

Fig. 4. x-axis: MS-BS distance; the unit is taken to be the quantization step $q = 553.8$ m. y-axis: measured TA