A UNIFIED APPROACH FOR HYBRID SOURCE LOCALIZATION BASED ON RANGES AND VIDEO

Beatriz Quintino Ferreira, João Gomes and João P. Costeira

Institute for Systems and Robotics, Instituto Superior Técnico, Universidade de Lisboa, Portugal
beatriz.quintino@tecnico.ulisboa.pt, {jpg, jpc}@isr.ist.utl.pt

ABSTRACT

This paper presents a hybrid method for single-source localization in wireless sensor networks, fusing noisy range measurements with angular information extracted from video. Although recent works found in the literature explore hybrid schemes, these include several cumbersome assumptions. We develop and test, both numerically and experimentally, a hybrid localization algorithm which surpasses the limitations of previous fusing approaches. The proposed method (FLORIS) is based on a nonconvex least-squares joint formulation, for which a tight convex relaxation is applied to obtain a semidefinite program. Numerical simulations show that FLORIS has comparable performance to state-of-the-art methods, even outperforming them in some scenarios. Real experiments show that FLORIS is feasible in practical application scenarios, achieving very good accuracy and robustness. Importantly, coverage requirements for the infrastructure in a given area are more flexible than resorting to a single type of sensor, which may simplify practical deployments.

***Index Terms*—** Hybrid single-source localization, convex relaxation, semidefinite programming, ranges, video

1. INTRODUCTION

The "where am I" problem has always been a key issue in the field of technology, both for human mobility and for robots/autonomous vehicles. Currently, the most popular localization system is the Global Positioning System (GPS). Nonetheless, there are several situations, such as indoors or underwater environments, in which GPS is not available and where location awareness will soon become an essential feature. These environments pose challenges such as strong multi-path/non line-of-sight propagation, diffractions or interferences, which lead to over-meter accuracy for the majority of existing systems. Such accuracy may be insufficient for numerous applications, and [1] claims that the key to overcome this issue lies on exploring hybrid schemes.

Focusing on indoor environments, most of the proposed localization systems use only one type of measurement. Yet, wireless sensor networks (WSN) are becoming ubiquitous and thus it is commonplace to find different sensors (e.g. Wi-Fi, Bluetooth, mobile cameras) inter-connected in the same space, the so-called Internet of Things. This work addresses the use of distances (obtained acoustically or with electromagnetic signals) and angular information (gathered by video cameras) to localize a target. More specifically, in WSN localization, range information can be measured from travel times [2], or inferred from received power [3], and usually produces robust results for ranges up to about 10 meters. On the other hand, orientation (Angle of Arrival) [4] and distance information retrieved by video is more reliable at short ranges. Therefore the complementary strengths of these techniques make them extremely appealing to be used in synergy, paving the way to more accurate localization.

Recently, attempts to fuse these two types of information have been presented in [4, 5], but these methods impose severe limitations; the one in [4] is specific for 2D, whereas [5] assumes that the range and visual anchors overlap. Our goal is to overcome the limitations of [4, 5] by deriving a novel formulation based on a *single* (centralized) optimization problem that jointly accounts for range and bearing data obtained from *arbitrarily placed* heterogeneous sensors in *2D or 3D*[1]. Using semidefinite relaxation (SDR) techniques [6] we obtain a convex problem that can be efficiently and globally solved by general-purpose software in one step.

SDR techniques have been successfully used before in range-based localization to obtain high-quality approximations to the maximum of the non-convex likelihood function under Gaussian noise [7, 8]. The approach that we take for the hybrid case builds upon our SLNN algorithm for range-only measurements [8], adopting related reformulations and relaxation techniques for a modified cost function that includes additional least-squares terms to account for angle measurements. SLNN's high precision and robustness is thus expected to carry over to the method proposed here. We include as one of the benchmarks another optimization-based

This research was partially supported by Fundação para a Ciência e a Tecnologia (FCT) through project PEst-OE/EEI/LA0009/2013 and EU FP7 project MORPH (grant agreement no. 288704)

[1]The approach is actually valid for an ambient space of arbitrary dimension.

method termed Squared Ranges Least-Squares (SR-LS) [9], which does not belong to the SDR class but provides an interesting tradeoff between good precision and very low computational complexity.

Our novel hybrid method for single-source localization, termed FLORIS (Fused LOcalization using Ranges and Incident Streaks), is fully tested in simulation and in real experiments with very encouraging results. In fact, we show that FLORIS outperforms the other benchmarks in some scenarios, particularly when the measurements are quite noisy. By taking advantage of hybrid measurements in our fully unified framework (as opposed to alternating between range-based and bearing-based localization in some ad-hoc schemes) a source may be localized even in extreme cases where the number of available ranges or bearings, taken independently, is insufficient to determine the position unambiguously.

Below, $(\cdot)^T$ denotes the transpose operator, $\mathbf{I}_n$ is the identity matrix of size $n \times n$, and $\otimes$ represents the Kronecker product.

2. PROBLEM FORMULATION: HYBRID SOURCE LOCALIZATION (FLORIS)

Let $x \in \mathbb{R}^n$ be the target position to be estimated based on a set of m known reference points (anchors) $a_i \in \mathbb{R}^n,\ i = 1, ..., m$. Of these, the ones whose indices belonging to set $\mathcal{R}$ provide range measurements to the source, $d_i = \|x - a_i\| + w_i$, where w_i denotes a zero-mean Gaussian noise term with standard deviation σ, whereas those with indices in $\mathcal{T}$ measure bearings. Each bearing, u_i, is modeled as a von Mises-Fisher random variable centered around the true direction $\frac{x - a_i}{\|x - a_i\|}$ with concentration parameter κ. We propose estimating the source position by minimizing the joint cost function

$$f(x) = \sum_{i \in \mathcal{R}} D^2(x, C_i) + \sum_{i \in \mathcal{T}} D^2(x, L_i), \quad (1)$$

where $D(x, C_i)$ denotes the distance from point x to the circle C_i centered at anchor a_i with radius d_i. Similarly, $D(x, L_i)$ denotes the distance from x to the line L_i that originates in a_i with orientation u_i.

The intuitive idea behind (1) is that this formulation attempts to balance, on the one hand, the distances of the target position estimate relative to the circles centered at the wireless anchors with radii d_i and, on the other hand, the distances to the lines originating at the visual anchors with orientation u_i (see Figure 1). Under i.i.d. Gaussian noise the first term in (1) is the likelihood of range measurements, but no such interpretation can be given for the second term.

While the squared distances in (1) are explicitly given by $D^2(x, C_i) = (\|x - a_i\| - d_i)^2$ and $D^2(x, L_i) = (x - a_i)^T(\mathbf{I_n} - u_i u_i^T)(x - a_i)$ [10], we pursue the following alternative parameterization for minimizing (1), inspired in [8]

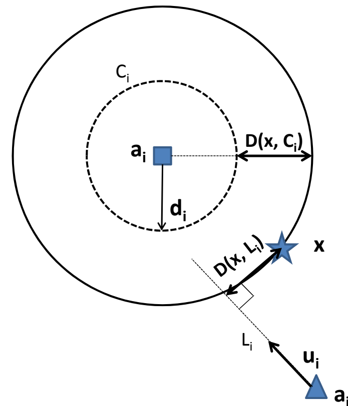

Fig. 1. Geometric representation of terms in the joint cost function (1)

$$\begin{aligned} &\underset{x, y_i, \theta_i, t_i}{\text{minimize}} && \sum_{i=1}^{m} \|x - y_i\|^2 \\ &\text{subject to} && y_i = a_i + d_i\theta_i, \quad \|\theta_i\| = 1, \quad i \in \mathcal{R}, \\ &&& y_i = a_i + u_i t_i, \quad t_i \in \mathbb{R}^+, \quad i \in \mathcal{T}. \end{aligned} \quad (2)$$

The first and second sets of constraints ensure that auxiliary variables y_i are located on the circles C_i or lines L_i, respectively. Problem (2) seeks the best possible match between these and x. Given all y_i, this is a standard least-squares problem whose optimal solution for x is just the center of mass of the constellation $x = \frac{1}{m}\sum_i y_i$, and this can be substituted back in the cost function of (2) to yield $\mathbf{y}^T\mathbf{J}\mathbf{y}$, where $\mathbf{y}$ is a vector of size $mn \times 1$ that stacks $y_1, \ldots, y_m$, and $\mathbf{J}$ is the projector on the orthogonal complement of $\mathbf{1}_m \otimes \mathbf{I}_n = [\underbrace{\mathbf{I}_n \ \ldots \ \mathbf{I}_n}_{m}]^T$. Compactly, we write

$$\mathbf{y} = \mathbf{a} + \mathbf{R}\begin{bmatrix}\boldsymbol{\theta} \\ \mathbf{t}\end{bmatrix}, \quad (3)$$

where $\boldsymbol{\theta}$ stacks the unit vectors $\theta_i,\ i \in \mathcal{R}$, and $\mathbf{t}$ stacks the scaling factors $t_i, i \in \mathcal{T}$. Matrix $\mathbf{R}$, of size $mn \times (n|\mathcal{R}| + |\mathcal{T}|)$, is block diagonal-like, built from $d_i\mathbf{I}_n,\ i \in \mathcal{R}$ and $u_i,\ i \in \mathcal{T}$[2]. Problem (2) is thus reformulated as

$$\begin{aligned} &\underset{\boldsymbol{\theta}, \mathbf{t}}{\text{minimize}} && \left(\mathbf{a} + \mathbf{R}\begin{bmatrix}\boldsymbol{\theta} \\ \mathbf{t}\end{bmatrix}\right)^T \mathbf{J} \left(\mathbf{a} + \mathbf{R}\begin{bmatrix}\boldsymbol{\theta} \\ \mathbf{t}\end{bmatrix}\right) \\ &\text{subject to} && \|\theta_i\| = 1,\ i \in \mathcal{R}, \qquad t_i \geq 0,\ i \in \mathcal{T}. \end{aligned} \quad (4)$$

Cost function (4) may be written as a quadratic form in $\begin{bmatrix}\boldsymbol{\theta}^T & \mathbf{t}^T & 1\end{bmatrix}^T$, and expressed using trace as

$$\operatorname{tr}\Big(\underbrace{\begin{bmatrix}\mathbf{R^T J R} & \mathbf{R^T J a} \\ \mathbf{a^T J R} & \mathbf{a^T J a}\end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix}\boldsymbol{\theta} \\ \mathbf{t} \\ 1\end{bmatrix}\begin{bmatrix}\boldsymbol{\theta}^T & \mathbf{t}^T & 1\end{bmatrix}}_{\mathbf{W}}\Big). \quad (5)$$

We now redefine the optimization variable as matrix $\mathbf{W}$ defined above, and rewrite the problem as

[2] Note that matrix $\mathbf{R}$ is not square, as each bearing u_i adds n rows but a single column to it. In spite of this nonstandard detail, its structure should be clear from the above description.

$$\begin{array}{lll} \underset{\mathbf{W}}{\text{minimize}} & \text{tr}(\mathbf{MW}) & \\ \text{subject to} & \mathbf{W} \succeq 0, \quad \text{rank}(\mathbf{W}) = 1 & \\ & \text{tr}(\mathbf{W}_{i,i}) = 1, \quad i \in \mathcal{R} & (6) \\ & \mathbf{W}_{i,nm+1} \geq 0, \quad i \in \mathcal{T} & \\ & \mathbf{W}_{nm+1,nm+1} = 1. & \end{array}$$

The third constraint, where $\mathbf{W}_{i,i}$ denotes the submatrix of $\mathbf{W}$ comprising the rows/columns that pertain to θ_i in (3), encodes $\|\theta_i\| = 1$. The fourth constraint, where $\mathbf{W}_{i,nm+1}$ denotes the subvector of $\mathbf{W}$ comprising the row pertaining to t_i in the rightmost column, encodes $t_i \geq 0$.

Finally, we drop the rank constraint in (6) to obtain the relaxed SDP. Vectors $\boldsymbol{\theta}$ and $\mathbf{t}$ are obtained by SVD factorization of the solution $\mathbf{W}$ or directly from its rightmost column (or bottom row), from which the y_i are computed by (3) and the source position estimated as the average of these m points.

Calibration: From the existence of two uncoupled types of sensors (range and visual), which contrasts with [4, 11], emerged the need to calibrate both networks, as a precondition to perform localization. Range measurements obtained acoustically will lead to a target position estimation in the coordinate system defined by the acoustic anchors. On the other hand, the orientation of the camera, relative to the identified visual features, is determined in a different coordinate system defined by such features. Hence, we should define a global frame and express both classes of measurements in this common coordinate system. Due to space constraints we omit the discussion of the calibration procedure, and assume throughout that suitable translations/rotations have been determined.

3. SIMULATION AND EXPERIMENTAL RESULTS

In this section we characterize the performance of FLORIS in simulation and in real experiments, benchmarking against SR-LS [9] and SLNN [8]. The latter were chosen based on the assessment of [8] which showed that SLNN has higher accuracy in 3D than previously proposed optimization-based methods, while SR-LS is somewhat less accurate but faster.

3.1. Simulation results

The following experiments were run using *MATLAB R2013a* and the general-purpose SDP solver CVX/SDPT3. Networks of acoustic and visual anchors were randomly generated in a [0, 5] × [0, 5] × [0, 5] cube. Gaussian noise was added to the distance according to $d = d_0(1 + w)$, where d_0 is the ideal (noiseless) range measurement and $w \sim \mathcal{N}(0, \sigma^2)$ is a zero-mean Gaussian random variable with standard deviation (noise factor) σ. Similarly, the orientation follows a von Mises-Fisher distribution $u \sim v\mathcal{MF}(u_0, \kappa)$, with mean direction u_0 denoting the ideal bearing and concentration parameter κ. Under this model range errors tend to increase

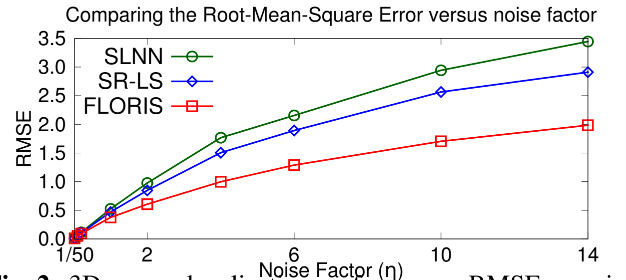

Fig. 2. 3D source localization performance; RMSE *vs* noise factor, for 6 acoustic and 4 visual anchors

for longer distances. A baseline scenario was established with reference values $\sigma_{ref} = 0.05$, $\kappa_{ref} = 2800$ that approximately reflects the dispersion of measurements in the experimental setup. To evaluate the performance as a function of noise power these were jointly scaled for other noise factors η as $\sigma = \eta\sigma_{ref}$, $\kappa = \frac{1}{\eta^2}\kappa_{ref}$. To assess the performance both the position estimation error and the rank of matrix $\mathbf{W}$ were studied. The rank is used to assess the formulation in (6) and its relaxation, since when rank$(\mathbf{W}) = 1$, the relaxed solution found holds the optimal value for the original non-relaxed problem. The accuracy is evaluated computing the Root-Mean-Square Error (RMSE), for every set of MC Monte Carlo runs, defined as $\sqrt{\frac{1}{MC}\sum_{i=1}^{MC}\|x_i - \hat{x}_i\|^2}$, where x_i and $\hat{x}_i$ are the true and the estimated source positions for the i^{th} run.

Example 1: Performing 1000 Monte Carlo runs for each noise factor value, for randomly generated configurations of 6 acoustic and 4 visual anchors, the obtained relative frequency of rank$(\mathbf{W}) = 1$ is listed in Table 1. We conclude that the relaxation used in FLORIS is tight, hence the optimal solution is frequently found, even for very high measurement noise.

η	1/50	1/10	1/5	1	2	4
Rank-1 (%)	99.9	97.4	96.6	84.8	85.7	82.8

Table 1. Percentage of rank-1 solutions for different 6+4 network configurations and noise factors

Example 2: A comparison among the three methods is performed in Figure 2 for several values of noise factors $\eta \in [\frac{1}{50}, 14]$, and for 1000 randomly generated network configurations, comprising 6 acoustic and 4 visual anchors. The figure shows significant improvement in accuracy by adding angular information, such that FLORIS consistently outperforms the other two (range-only) state-of-the-art methods. It can be remarked that the comparison is unfair, as SLNN and SR-LS operate with less information (6 acoustic anchors alone) than the hybrid method (which adds 4 additional visual anchors). A more equitable situation is discussed next.

Example 3: Figure 3 depicts simulation results when all algorithms use the same number of anchors. Networks of 8 acoustic anchors were generated to test SLNN and SR-LS, 4 of which were randomly converted to visual ones for the hybrid approach. Under these conditions, the performance of FLORIS is closer to that of other methods. In fact, for very

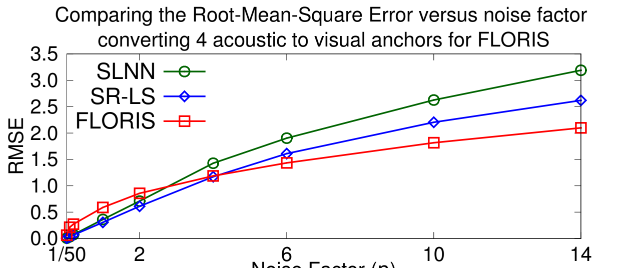

Fig. 3. 3D localization performance, RMSE *vs* noise factor for 8 acoustic or 4 acoustic + 4 visual anchors

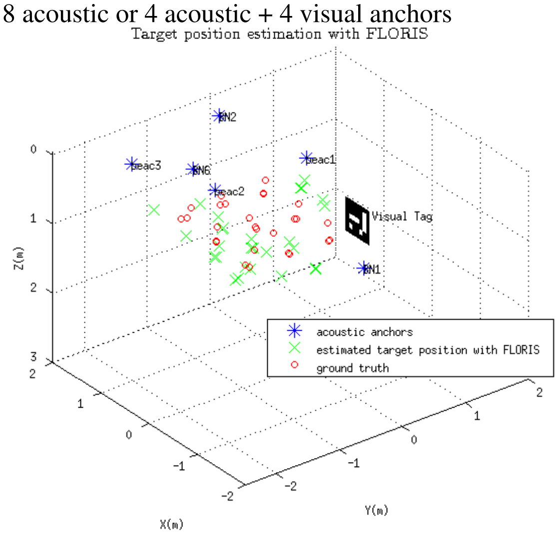

Fig. 4. Target position estimations given by the hybrid algorithm versus the ground-truth positions

low noise factors it is slightly outperformed by SLNN and SR-LS. On the other hand, the proposed method seems to be consistently more robust for noisier measurements. This is an important property of FLORIS, as measurements in practical scenarios tend to be quite noisy. Running times (on the order of 1 sec) are similar to SLNN's.

3.2. Experimental results

An experimental set-up was developed to test FLORIS. This consisted on Cricket [12] beacon nodes as acoustic anchors and ARUCO [13] augmented reality tags as visual anchors. The target, comprising a video camera rigidly coupled to a Cricket listener node, could roam in a covered volume of about 50 m^3. Several practical issues had to be overcome, including tuning the directionality of Crickets, calibrating their range measurements individually, and setting up transformations to translate between the coordinate systems of visual and acoustic nodes (see the end of Section 2).

Several datasets (of range and orientation measurements) were acquired. Figure 4 shows the ground truth and the target positions estimated by FLORIS during a walk through the installed set-up. It can be observed that the estimated positions are, globally, very close to the ground truth.

Table 2 compares the RMSE obtained for the proposed algorithm with SR-LS [9] and SLNN [8], for a particular data set. For this specific case, it can be observed that FLORIS does improve upon existing methods, corroborating the good performance already found in simulation. This is a very encouraging result, showing that although several practical limitations were found in the deployment of the experimental set-up they were successfully overcome.

	SR-LS [9]	SLNN [8]	FLORIS
RMSE(m)	0.1738	0.1973	0.1609

Table 2. 3D source localization performance comparison, for an experimental data set obtained with Cricket and ARUCO anchor nodes

Overall, we believe that our experimental results validate FLORIS as a practically relevant algorithm with appealing accuracy and robustness properties. We stress that FLORIS is potentially more flexible than other methods operating on a single type of sensed variable. More specifically, the hybrid approach is able to successfully localize a target in situations where other algorithms are not. For example, in 3D, when information from only 3 acoustic anchors is received, it is not possible to estimate a correct position. Yet, if an object (or a tag in our current implementation) is recognized FLORIS produces a valid estimation. Furthermore, although FLORIS experiences a well-known deterioration in performance as the source moves outside the convex hull spanned by the anchors, we have observed that the degradation is more progressive and graceful than in range-only algorithms. So far, the evidence for this effect remains only anecdotal, and a more careful characterization should be undertaken.

4. CONCLUSIONS AND FUTURE WORK

This work addresses the growing need for solutions and applications for large-scale heterogeneous sensor networks, taking advantage of distinct sensing devices. In particular, we devised a new approach for centralized localization based on nonlinear least squares that seamlessly fuses range and angular measurements, as well as a tight SDP relaxation that provides an efficient solution method using a generic convex solver. FLORIS was numerically and experimentally validated, showing that it can provide accurate position estimates. In numerical results FLORIS achieves higher accuracy than state-of-the-art methods, specially for high noise scenarios. Findings from a real indoor deployment are promising, showing better accuracy of estimates than the benchmarks, including in cases that cannot be tackled by previous methods using a single type of sensor. Noisy acoustic range measurements are an important practical concern, and alternatives (e.g. using low-cost UWB nodes) are being assessed. The procedure to translate between coordinate systems for the various types of nodes also requires improvement. Longer-term future work will explore the replacement of visual anchors (tags) by detection/recognition of more general artificial objects.

5. REFERENCES

[1] D. Schneider, "You Are Here," *IEEE Spectrum*, vol. 50, no. 12, pp. 34–39, Dec. 2013.

[2] J. Bachrach and C. Taylor, *Localization in Sensor Networks, in Handbook of Sensor Networks: Algorithms and Architectures*, John Wiley & Sons, 2005.

[3] P. Oguz-Ekim, J. Gomes, and P. Oliveira, "RSS Based Cooperative Sensor Network Localization with Unknown Transmit Power," in *Proceedings of Signal Processing and Communications Applications Conference (SIU), 2013 21st*. 2013, pp. 1–4, IEEE.

[4] P. Biswas, H. Aghajan, and Y. Ye, "Integration of Angle of Arrival Information for Multimodal Sensor Network Localization using Semidefinite Programming," in *Proceedings of 39th Asilomar Conference on Signals, Systems and Computers*, 2005.

[5] M. Crocco, A. Del Bue, and V. Murino, "A Bilinear Approach to the Position Self-Calibration of Multiple Sensors," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 660–673, Feb. 2012.

[6] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[7] K. Cheung, W. Ma, and H. So, "Accurate approximation algorithm for TOA based maximum likelihood mobile location using semidefinite programming," in *Proceedings of IEEE Int. Conf. Acoust., Speech, Signal Process (ICASSP'04)*, Montreal, Canada, 2004, pp. 145–148, IEEE.

[8] P. Oguz-Ekim, J. Gomes, J. Xavier, M. Stosic, and P. Oliveira, "An Angular Approach for Range-Based Approximate Maximum Likelihood Source Localization Through Convex Relaxation," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3951–3964, July 2014.

[9] A. Beck and P. Stoica, "Exact and Approximate Solutions of Source Localization Problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770–1778, May 2008.

[10] J. P. Ballantine and A. R. Jerbert, "Distance from a line, or plane, to a point," *The American Mathematical Monthly*, vol. 59, no. 4, pp. 242–243, Apr. 1952.

[11] M. Crocco, A. Del Bue, I. Barbosa, and V. Murino, "A Closed Form Solution for the Self-Calibration of Heterogeneous Sensors," in *Procedings of the British Machine Vision Conference 2012*, 2012.

[12] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan, "The Cricket Location-Support System," in *The Proceedings of the sixth ACM International Conference on Mobile Computing and Networking (ACM Mobicom 2000)*. 2000, pp. 32–43, ACM.

[13] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280 – 2292, 2014.