Geometry calibration of distributed audio-visual sensor networks

Florian Jacob

Department of Communications Engineering

01.07.2014
1 Introduction & motivation
2 Acoustic calibration methods
3 Experimental results (1)
4 Audio-visual calibration
5 Experimental results (2)
6 Summary & future research
Motivation

What does “geometry calibration” mean?

- Determine positions and orientations of sensor nodes from measurements

Application areas of spatially distributed microphone arrays

- Speaker localization & tracking
- Signal enhancement (i.e. source separation)
- Spatial signal processing (i.e. acoustic beamforming)

⇒ Geometry calibration required

Goal

- Automatic calibration instead of exhausting, time-consuming, manual process!
Motivation

Classification of existing approaches, according to ...

- Available observations:
  - Time of flight (ToF)
  - Time difference of arrival (TDoA)
  - Direction of arrival (DoA)

- Used sound sources:
  - Pilot sequences
  - Special playback apertures
  - Natural sound sources (i.e. speech signal)

Our approach

- DoA based calibration ⇒ Relaxed synchronization requirements
- Speech signal ⇒ Low requirements for input data
Common approach [KL08]
Calibration method

- Objective function \( f(\Omega) \) consists of geometric relation for each microphone array and observation:

\[
f_{i,t}(\Omega) = (x_i + y_i \tan(\theta_i) - a_t) \tan(\varphi_{i,t}) - y_i - (a_t + b_t \tan(\varphi_{i,t}) + x_i) \tan(\theta_i) + b_t
\]

- Requires at least \( N \geq \frac{3(K-1)}{K-2} \) independent observations for \( K \) sensors
- No closed form solution \( \Rightarrow \) numerical solution by Newton’s method:

\[
\Omega_{\kappa+1} = \Omega_{\kappa} - J^{-1}(\Omega_{\kappa}) \cdot f(\Omega_{\kappa})
\]

- Arbitrary microphone array defines origin of coordinate system
Numerical issues

- Numerically unstable for angles in the vicinity of ±\( \frac{\pi}{2} \)
- Reason: Singularities of tangent function

Solution

- Reformulate geometric relation

\[
\mathbf{f}_{i,t}^{\text{cos}}(\Omega) = \mathbf{v}_{i,t}^{\text{T}} \mathbf{R}(-\theta_j + \frac{\pi}{2}) \mathbf{v}_{i,t} \cdot \mathbf{a}_t - x_i - b_t - y_i.
\]
Analysis & improvements

Solution invariant to sensor orientation

- Two local minima for each sensor orientation
- Reason: Orthogonality is not unique

\[ f^\text{cos}_{i,t}(\Omega) = \begin{bmatrix} \cos(\varphi_i,t) & \sin(\varphi_i,t) \end{bmatrix}^T \begin{bmatrix} \sin(\theta_j) & -\cos(\theta_j) \\ \cos(\theta_j) & \sin(\theta_j) \end{bmatrix} \begin{bmatrix} a_t - x_i \\ b_t - y_i \end{bmatrix} \]

\[ g^2_{i,t}(\Omega) = \begin{bmatrix} \cos(\varphi_i,t) & \sin(\varphi_i,t) \end{bmatrix}^T \begin{bmatrix} \cos(\theta_j) & \sin(\theta_j) \\ -\sin(\theta_j) & \cos(\theta_j) \end{bmatrix} \begin{bmatrix} a_t - x_i \\ b_t - y_i \end{bmatrix} - \sqrt{(\tilde{v}_{i,t}^T \tilde{v}_{i,t})} \]

Proposed improvement

- Avoids the existence of minima, that correspond to wrong sensor orientations!
Summary

- Geometric relation is scale invariant $\Rightarrow$ only relative calibration possible
  - ✗ Additional scale information for absolute geometry required
- Numerically unstable for angles in the vicinity of $\pm \frac{\pi}{2}$
  - ✔ Reformulation of equation system
- Sensor configuration rotation invariant
  - ✔ Reformulation of equation system
- Newton’s method sensitive to initial values (local minima or divergence)
  - ✔ Monte-Carlo based initialization
- Distorted DoA measurements cause problems
  - ✔ Random Sample Consensus Algorithm (RANSAC)
Random Sample Consensus (RANSAC) Algorithm [FB81]

**Overview**

- **Goal:** Robust estimation of model parameters in presence of outlier measurements
- **Application areas:**
  - Cartography
  - Image processing

Available observations: \( \Psi = \{\psi_1, .., \psi_m\} \)

\[ \Omega := n \text{ observations chosen from } \Psi \]

compute model parameters for \( \Omega \)

consensus := \( \{\psi_i| \text{fit to model, } \forall i\} \)

\#(consensus) < threshold \quad \#(consensus) > threshold
Calibration framework

- Continuous DoA estimation delivers large number of observations
- Geometry calibration based on Newton’s method is computationally expensive
Experimental results (1)

Simulated environment

- Randomly walking speaker
- Sound propagation simulated by Image method
- Reverberation times: 0 ms – 500 ms
- DoA estimation: Filter-and-sum beamformer (FSB)

![Diagram showing trajectory and sensor arrays](image-url)
Audio-visual calibration

Goal

• Joint audio-visual speaker tracking with a self-calibrating sensor system

Assumptions

• Reference: Visual sensor network
• Acoustic sensor network calibrated

Calibration approach

• Separate speaker tracking for acoustic and visual sensor network
• Map acoustic trajectory onto visual trajectory
• Mapping parameters required for joint calibration
Coordinate mapping problem

Mapping between acoustic and visual location estimates

- Acoustic location estimates \( m_i \)
- Visual location estimates \( c_i \)

\[ c_i = sRm_i + t \]

⇒ Estimate Rigid Body Transformation (RBT) parameters:
  - Rotation \( R \)
  - Translation \( t \)
  - Scale \( s \)

Least squares objective function

\[
\langle R^*, t^*, s^* \rangle = \arg\min_{R,t,s} \frac{1}{N} \sum_{i=0}^{N-1} \|sRm_i + t - c_i\|^2
\]

Solution

- Common approach: Singular Value Decomposition based [Cha95]
- Proposed: computation in Shape domain
Available information

- Visual DoA: Histogram of Orientated Gradient (HOG) detection of head and shoulder
- Acoustic DoA: Filter-and-sum beamformer (FSB)
Simulation setup

Scenario
- 4 virtual cameras
- 4 microphone arrays (2-elements)
- 2 sensor setups
- 5 random trajectories for each setup

Data generation
- Visual DoA: HMM based error model trained on AV16.3
- Acoustic signal based on Image method
Simulation results

Sensor positioning and orientation error

![Graph showing sensor positioning and orientation error]

Speaker localization error

<table>
<thead>
<tr>
<th>$T_{60}$ [ms]</th>
<th>0</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>audio</td>
<td>0.09</td>
<td>0.18</td>
<td>0.32</td>
<td>0.43</td>
<td>0.55</td>
<td>0.66</td>
</tr>
<tr>
<td>video</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>combined</td>
<td>0.08</td>
<td>0.15</td>
<td>0.22</td>
<td>0.26</td>
<td>0.30</td>
<td>0.33</td>
</tr>
<tr>
<td>Joint</td>
<td>0.07</td>
<td>0.13</td>
<td>0.16</td>
<td>0.17</td>
<td>0.18</td>
<td>0.18</td>
</tr>
</tbody>
</table>
Summary & future research

Acoustic self-calibration

- Acoustic calibration algorithm using reverberant speech input
- Reformulation of the calibration problem
  - Numerical issues solved
  - Rotation invariance avoided
- Calibration error significantly reduced by RANSAC

Joint calibration

- Localization error of visual sensor network decreased by self-calibrating acoustic sensor network

Future research

- Scale factor estimation
- Joint calibration of acoustic and visual sensor network
- Evaluation in real scenarios
Challis, J. H.:
A procedure for determining rigid body transformation parameters.
In: Journal of Biomechanics (1995)

Fischler, M.; Bolles, R.:
Random Sample Consensus: a paradigm for model fitting with applications to image analysis
and automated cartography.
In: Communications of the ACM (1981)

Kemper, J.; Linde, H.:
Challenges of passive infrared indoor localization.
In: 5th Workshop on Positioning, Navigation and Communication, 2008
Thank you for your attention!

Questions?

Florian Jacob
University of Paderborn
Department of Communications Engineering
jacob@nt.uni-paderborn.de
nt.uni-paderborn.de