Active Sound Localization in a Symmetric Environment

Regular Paper

Jordan Brindza1,*, Ashleigh Thomas1, Spencer Lee1, William McDermid1, Yizheng He1 and Daniel D. Lee1

1 General Robotics Automation, Sensing and Perception (GRASP) Laboratory, University of Pennsylvania, Philadelphia, PA, USA

* Corresponding author E-mail: brindza@seas.upenn.edu

Received 14 Sep 2012; Accepted 22 Apr 2013

DOI: 10.5772/56574

© 2013 Brindza et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Localization for humanoid robots becomes difficult when events that disrupt robot positioning information occur. This holds especially true in symmetric environments because landmark data may not be sufficient to determine orientation. We propose a system of localizing humanoid robots in a known, symmetric environment using a Rao-Blackwellized particle filter and a sound localization system. This system was used in the RoboCup Standard Platform League, and has been found to reduce the amount of own-goals scored as compared with the previously used localization system without sound.

Keywords Auditory Localization, Bipedal Locomotion, Binaural Signal Processing, Autonomous Soccer

1. Introduction

The RoboCup Standard Platform League (SPL) is an international competition for autonomous, humanoid soccer playing robots [1]. In the SPL, universities from around the world program teams of Aldebaran Nao robots to compete in autonomous 4 versus 4 soccer matches. The RoboCup competition is designed to promote research in the fields of humanoid robotics, computer vision, machine learning, localization and multi-agent collaboration.

The Aldebaran Nao (Figure 1) is a 57.3 cm tall humanoid robot with 21 degrees of freedom and powered by an Intel ATOM processor and 1 GB of RAM. The Nao has two front-facing, non-overlapping 1.22 MP color cameras, two microphones and two loudspeakers mounted on its head. It also contains two 1-axis gyroscopes and one 3-axis accelerometer [2].

Although RoboCup is concentrated on soccer playing humanoids, many advancements are directly applicable to robotics research in general. Robots of all kinds need to locomote, see, localize and cooperate. The aim of RoboCup is to advance these fields more rapidly through a competitive environment. Each year, the rules are changed to make the competition more challenging and realistic.

Traditionally, RoboCup matches were played on a field with two unique goals, one blue and one yellow. The differently coloured goals disambiguated the otherwise symmetric environment. New, more challenging rules
were recently introduced that changed the uniquely coloured goals to two identically coloured goals, just like is the case on real soccer fields. This change created a completely symmetric environment for the robots to localize in and a new set of challenges in determining which goal to defend and which goal to attack.

Figure 1. The Aldebaran Nao humanoid robot.

Even in an asymmetric environment, the physical nature of a soccer match makes reliable localization difficult. The game is played with humanoid robots that have physical contact with one another and are prone to falling. It is difficult to keep track of which direction the robot falls or how far the robot moves during a fall. After falling, the robot loses its heading completely and can move a full meter by the time it has finished standing up. The robots also have to handle those situations where they are “kidnapped” and removed from play after committing a foul. A robot is returned to its half of the field and positioned somewhere on the side-line furthest from the ball. If the environment is asymmetrical, it is possible to re-localize through the use of visual landmarks alone. In a symmetrical field, however, these visual clues are non-existent, making re-localization a very difficult task.

Symmetric environments are common in the developed world; building and room layouts are many times symmetric, and streets and their intersections have 2-fold and 4-fold symmetry, respectively. One can easily imagine giving a team of robots a building blueprint or a road map and expecting them to navigate successfully, despite un-mapped obstacles or obstructions such a debris, furniture, or a snag in a rug.

Sharing information between teammates over a wireless network is a common strategy for overcoming the symmetric field problem, but robots must be able to localize even when wireless communication is faulty. Other approaches include the detection of sparse visual cues in the environment surrounding the field, but these cues are often unreliable because the area around the field is not controlled and can be constantly changing as spectators move around the field.

We propose a solution to the symmetric field problem by using a sound emitted by the goalkeeper to inform the other robots on the team as to which goal they are defending. We attempt to disambiguate identical goals by using the goalkeeper as an artificial landmark. By “listening” to the sound emitted by the goalkeeper, players can determine the direction of the goalkeeper and, therefore, the direction of the goal they are defending.

2. Related Work

There has been extensive research by RoboCup teams about localizing in an asymmetric environment with two unique goals. There are two methods that are used by most teams: Kalman filters and particle filters. The UT Austin RoboCup team uses multiple Kalman filters [3]; every time there is an ambiguous detection, new filters are created for each possible outcome. Certainties associated with each filter are used to determine when a model becomes so improbable that it should be discarded. Particle filters have also been adapted to function efficiently for humanoid robots [4]. This filter uses a set of particles to represent multiple hypotheses for the current robot state. These particles are updated independently based upon the robot’s motion and sensor input. The particles are then weighted based on the sensor readings to determine the most likely hypothesis.

There has been work on adjusting particle filters to better deal with symmetric environments, but many algorithms only localize up to symmetry [5]. This means that the filter will converge on any one of many identical locations, as shown in Figure 4. In the context of robot soccer, this can result in scoring an own goal.

Instead of using the particle with the highest certainty to determine the robot’s position, some algorithms wait for particles to reach a consensus and then use an average of the synchronized particles. In [6], particles are grouped into clusters by their headings, whereby particles with the same orientation are grouped together. If a cluster has enough particles, the average of the particles in the cluster is deemed to be the correct position of the robot. This approach was shown to be more accurate than the traditional particle filter implementation. However, it can still only localize up to symmetry.

A popular method for dealing with the symmetric field problem involves sharing localization information between teammates over a wireless network [7,8]. These approaches maintain the typical individual localization systems for a symmetric field, and the differentiation of the sides of the field is determined by gameplay cues. These methods also make use of a shared ball model and side confidence values of teammates to self-correct when confidence is lost in position estimates. This confidence is
negatively affected by events such as collisions with other players or kidnappings. For an individual robot to regain confidence in self-localization, it refers to the localization information of teammates with high confidence values. The robot can re-localize by comparing the observed location of the ball with the shared ball model. However, when using a team ball position model, there is a risk of delocalizing the entire team due to the propagation of errors in ball position estimates. The principle drawback of these methods is that they rely on network communication between teammates. If the network is not reliable or if several robots are penalized, the individual robots will not be able to localize accurately.

Yet another proposed method has been to use external, visual landmarks to break the field symmetry. One of these proposed methods scans the environment for colour, building a colour histogram model [9]. This model, in conjunction with self-localization based on traditional particle filtering, is used to differentiate the sides of the field. The particle weights are adjusted according to the correlation between the currently perceived surroundings and the expected surroundings of the colour histogram model. In [10], a similar method is proposed which uses one-dimensional SURF features to learn the appearance of the area surrounding the field. A major weakness in these approaches is that the environment around the field is not fixed and can change drastically as people and objects in the area move around and consequently invalidate the learned appearance models. These methods will also fail if an environment has no symmetry-breaking features.

Our proposed solution to the symmetric field problem involves the use of sound emission and detection to determine the direction of the defending and attacking goals relative to a robot’s current position. Other works making use of sound localization include [11,12]. Previous research on this technique has focused on domestic robots, where background noise is not as much of a problem as during a soccer game.

The use of head-related transfer functions (HRTFs) is a popular method for sound localization [13]. The shape of the head and, to a lesser extent, the body of a robot affects audio signals as they reach the microphones of that robot.

An HRTF is a transfer function linking an unaffected audio signal to the signal sensed by the robot. Comparing sensory data to a predetermined signal using an HRTF yields enough information to determine the direction, elevation and distance to the signal’s source. Unfortunately, HRTFs are difficult to design and the associated signal source analysis is computationally costly. Additionally, a true HRTF is based not only upon the physical structure of the “listening” robot, but also upon the structure of the environment around it. Since environments are rarely static when dealing with mobile robots, it is impossible to design a truly accurate HRTF. Our method, which relies upon only the time difference between receiving a signal at each side of the robot’s head, can only determine the direction from which the signal originates, but this computation is much cheaper than one using an HRTF and it still gives the necessary orientation information.

3. Localization

We use a hybrid Rao-Blackwellized particle filter approach for localization. We track the Cartesian coordinates and the heading of each robot, as well as a probabilistic measure of uncertainty (see [14]). Our approach combines discrete Markov updates and Kalman filter updates to estimate the orientation and coordinate position, respectively. The accuracy of our pose estimate is represented by two-dimensional Gaussians. The estimation obtained with Kalman filtering is a product of two steps: the motion model update and the measurement update.

The motion model update - also referred to as the odometry update - utilizes the robot kinematics to update the particle filter as the robot walks around the field. Given the joint angles of the robot, forward kinematics are used to compute the location of the robot’s feet as it walks. The change in translation and rotation of the body of the robot are computed based on the position of the feet, as shown in Figure 2, and used to update the particle filter.

![Figure 2. Odometry update after the robot takes one step with the left foot.](image)

The measurement model refines this estimate using sensory inputs, such as vision-based landmark detection. The colour cameras are the main sensors used by the robot to perceive its environment. We use supervised learning to train a Gaussian mixture model for classifying the important environment colours: ball (orange), goals (yellow), ground (green) and the lines (white). The classifier is trained offline from logged images and stored on the robot.
The colour classifier is used to label images from the camera in real-time. Once the pixels in a picture are classified, they are segmented into connected components, after which landmarks like the ball and goalposts can be located. To detect lines and corners, a Hough transform is used to search for relevant line directions. Further checks are implemented to exclude lines that are above the ground plane or outside the field, and intersecting lines are merged to detect the corners of the field. Finally, the full 3D pose of the field landmarks can be computed because the real dimensions of the field landmarks are known.

Figure 3. Image processing results for a single frame. The figure shows the original image on the left overlaid with the object detections for the ball, goal posts and field lines. The right image is the image following colour classification.

The measurement model incorporates these field landmark detections to adjust and weight the particles in the filter. The measurement model works well for small corrections in the position and orientation of the robot pose, but it cannot break down the symmetry of the field. Figure 4 shows an example of the particle filter converging on two symmetrically identical poses on the field.

Figure 4. Example of particle filter convergence in a symmetric environment. The red triangle represents the actual robot position and the blue arrows represent the various particle hypotheses.

Under ideal circumstances - and if the robot’s starting position is known - the particle filter approach alone is enough to keep track of the correct robot pose. However, noise in the motion model, inevitable false positive detections of field landmarks, and falling down, will all eventually cause the robot to converge on a pose that is symmetrically opposite the true location. This is further complicated by kidnapping, which happens when the robots are penalized. Thus, a robust method for disambiguating the halves of the playing field is crucial to maintaining the correct localization of the robot. Without it, the robots risk scoring ‘own’ goals.

4. Sound Localization to Disambiguate Goals

Our solution to disambiguate the goals is to use the goalkeeper as a beacon so that the other robots on the team can determine the direction of the goal they are defending. We use the goalkeeper as a beacon because it is the player with the most robust knowledge of the field orientation and it remains in roughly the same location throughout the match. Since it stays well away from the centre of the field, falling down does not introduce enough error into its localization to cause ambiguity in the orientation of the field. It is therefore safe to assume that the goalkeeper will not lose track of the correct orientation of the field and will stay in front of its defended goal. Using vision to detect the goalkeeper is an attractive approach but suffers from a number of complications. There is very little - only the robot’s jersey number - to differentiate the goalkeeper from another teammate and it becomes increasingly difficult at large distances. In addition, there is rarely a clear line of sight to the goalkeeper during a match due to occlusions from other robots on the field. We therefore chose to use an auditory signal for localization.

While the goalkeeper is in front of its defending goal it transmits an audio signal through its loudspeakers at around 70 Db. The other robots on the field listen for this audio signal and use its time of arrival to their left and right microphones to estimate the relative direction of the goalkeeper. We use a pseudo-random audio signal that is generated offline and known by all of the robots on the team. The pseudo-random signal was chosen so as to avoid misdetections from background noise or from the opposing team.

Cross-correlation is used to detect the pseudo-random signal recorded by the microphones [15]. Let $f$ represent the raw audio samples from the microphones and $g$ represent the known pseudo-random signal that is being transmitted by the goalkeeper; then, the cross-correlation is:

$$ (f \ast g)(t) = \int_{-\infty}^{\infty} f^*(\tau) g(t + \tau) d\tau $$

where $f^*$ indicates the complex conjugate. The cross-correlation between the two signals can be computed efficiently using the Fast Fourier Transform (FFT) [16]. This is done independently for the left and right audio
signals. If the predetermined signal is detected in both the left and right microphones, the disparity is the time offset between these signals as shown in Figure 5.

Figure 5. Plot of the left (blue) and right (red) cross-correlation with the pseudo-random signal. The left and right channels are synthetic data generated from the pseudo-random signal with additional Gaussian white noise added.

This stereo disparity is used to determine the direction of an audio sound source [17,18]. The direction of the source relative to the microphone baseline, \( \theta \), can be found as follows:

\[
\sin \left( \frac{\pi}{2} - \theta \right) = \frac{v}{b} (t_1 - t_r) = \frac{v(n_l - n_r)}{bf}
\]

(2)

where \( v \) is set to the speed of sound, 340.29 meters per second, which we assume remains constant; \( b \) is set to 12.12 cm, the baseline of the stereo microphone pair; \( t_1 \) and \( t_r \) are the time of arrival at the left and right microphones, respectively, which can be represented in terms of the left \( (n_l) \) and right \( (n_r) \) audio samples divided by the sampling rate \( (f = 16,000 \, \text{Hz}) \).

Auditory epipolar geometry, unlike visual epipolar geometry, has an inherent ambiguity in the direction of the sound source. The location of a single detection of an auditory signal from a single source by the stereo pair of microphones can only be resolved up to a right cone [17]. This “cone of confusion” has a vertex at the midpoint between the two microphones, with the baseline forming the axis of symmetry as shown in Figure 6.

We can further simplify the problem because we are only interested in the direction of the sound in the horizontal ground plane. The height of the sound source above the ground is fixed by ensuring that the goalkeeper only transmits the auditory signal while it is standing upright. Therefore, the direction of any single sound source can be resolved to up to two ambiguous directions. The sound source can be located at the positive and negative angle from the baseline, as shown in Figure 7.

Figure 6. Visualization of the 3D cone of confusion resulting from a sound source located at the red marker.

To verify the robot’s ability to estimate the direction of a sound source from the audio disparity, we fixed the position and orientation of the listening robot. Then, we moved the robot transmitting the audio signal around the listening robot in 15° intervals and recorded the direction of the sound source estimated by the listening robot for 6 trials at every position. The experiments were repeated at distances of 1, 2, 4 and 5 meters, a sample of the typical range of distances between the goalkeeper and the other robots during a match. Figure 8 shows the data from the experiment. The measurements were limited from 0 to 90° because the robot symmetry and cone of confusion make the results from the remaining workspace redundant. Our limit on the range of the data has, however, skewed some results for the largest and smallest angles being measured. Taking data in the 0-90° range means that a reading of an angle slightly less than 0° or slightly more than 90° is recorded as slightly more than 0° or slightly less than 90°, respectively. As a result, the accuracy of our systems as measured by this test seems to degrade at the low and high end of our testing spectrum.

Figure 7. Visualization of the 2D directional ambiguity when using stereo disparity for a sound source located at the red marker.

It is possible to identify the true direction of a sound source given two or more detections while the microphones are in different orientations. In order to obtain the true direction of the sound source, the robot must hear the noise in two orientations that differ by an
angle greater than or equal to the audio detection discretization. This would require detections that are 10° apart in our case, due to the distance between the microphones and the audio sampling rate that we use (16,000 Hz). This is easily accomplished because the microphones are mounted on the robot’s head and can be rotated by either changing the head yaw or by having the robot turn by walking. The robots are constantly walking around the field to move into position or rotating their heads to look for the ball and field landmarks, so no additional routine is needed to acquire multiple audio detections at different orientations.

The audio localization is only capable of determining direction of the source from the audio correlation on the vertical axis at different distances between the robot and the sound source.

To keep track of the current goalkeeper position, we use a polar log-likelihood map oriented relative to the robot’s body. The map is discretized into 30° bins. For each successful audio correlation, the disparity is used to calculate the direction of the sound source and the two ambiguous directions in the filter are updated. The direction of the goalkeeper is determined as being known with a high degree of confidence when the global maximum direction hypothesis is significantly greater than any other local maximum in the filter. For our implementation, we considered the direction of the goalkeeper to be certain if it was at least two times larger than any other local maximums.

The audio localization is only capable of determining direction, and there is no robust way to extract the distance of the source. This is a problem because large translations of the listening robot can cause the sound filter results to become invalid. In the worst case, when the robot walks across the field and through the position of the goalkeeper, this will result in a change of up to 180°. In order to account for this, we add a large, discrete decay to the filter that is based on the robot’s translation in addition to the continuous time decay. The sound filter keeps track of the robot’s reference position - initialized to the starting position - and calculates the robot’s translation from odometry. After moving one meter from its reference position, a large decay is applied to the filter and the reference position is updated to the robot’s current position. Similarly, when the robot detects that it has been “kidnapped”, the sound filter and particle filter are both reset. The robot can use the force sensors on its feet to tell whether it is no longer in contact with ground and, therefore, has been kidnapped.

The sound localization system is only used as input to the particle filter and does not require any additional modifications. The sound source filter is continuously running in the background and is only used to label goal post detections from the vision system as either attacking or defending goals. The directions of all goal post detections are compared to the estimated direction of the goalkeeper found from the filter. If the directions are the same - up to a threshold - then the post can be labelled as a defending goal post instead of an unknown goal post. The particle filter handles this naturally by weighting the particles in the correct pose over those in the symmetric pose.

Finally, an important consideration is the amount of processing needed to run the algorithms. In the RoboCup SPL, the robots must run all their code on board, without using any external resources, and the audio detection and correlation must run together with the motion and image processing algorithms. This places stringent limits upon the amount of time and processing power that can be devoted to detecting the auditory signal. The motion algorithms controlling the walk engine must update at 100 frames per second (fps) to keep the robot stable and the image needs to update at up to 30 fps. The robot is, therefore, incapable of continuously running the audio correlation. We solved this resource constraint problem by gating the correlation computation using a dual-tone multi-frequency (DTMF) signal [19].

The goalkeeper is set to transmit a DTMF signal immediately before playing the pseudo-random audio signal. The robots listening for the goalkeeper signal only need to compute one FFT per second until the DTMF signal is detected. This greatly reduces the amount of processing required to detect and correlate the audio signal. A naïve implementation that attempts to compute the correlation continuously would require ~120 FFTs of a 512 sample signal per second. The sound filter performance will increase as the beacon signal frequency increases because it gives more chances to successfully correlate the signal and update the filter. However, the required processing power is dependent upon the signal frequency and so there is a trade-off between processing power and filter performance. We found that a signal frequency of 1 Hz provides a good balance between

Figure 8. The graph shows the actual direction of the sound source from the robot on the horizontal axis and the estimated direction of the source from the audio correlation on the vertical axis at different distances between the robot and the sound source.
processing power and filter performance. Assuming the goalkeeper transmits the signal once every second, gating the computation with a DTMF signal requires ~64 FFTs of 512 sample signals per second, a 47% decrease in computation time. We use an integer-based FFT implementation for computing the cross-correlation, which results in a very efficient computation. While running the cross-correlation takes less than 5% of the CPU on board the robot.

The main weakness of this approach is the reliance upon the proper localization of the goalkeeper. In the event of the delocalization of the goalkeeper, the error would propagate to the other team member. Additionally, we assume that the robots are in an environment that does not produce large echoes. From our experience, this has been true for soccer fields. Finally, we do not address the case where a hostile agent determines the signal we are using and plays it to confuse our localization system. It is against the RoboCup SPL rules to do this and, as a result, we do not consider this scenario.

5. Conclusion

In this paper we presented a technique for localizing in a symmetric environment with a multi-robot team. Our solution involves using auditory localization to disambiguate the direction on a symmetric soccer field. One robot, the goalkeeper, acts as a beacon that the other robots use to determine the direction of their defending goal. The robots use this information to disambiguate goal observations used as inputs to the particle filter. This approach allows us to correctly localize on the symmetric field even after being kidnapped or falling down. We have shown that the robots can robustly and efficiently localize a sound source using stereo audio disparity from a pair of head-mounted microphones. We successfully implemented and deployed this approach at the RoboCup 2012 Standard Platform League competition.

Our approach does not rely upon an undependable infrastructure, such as a wireless network, or on the uncontrollable field surroundings as previous solutions have. The main drawback of utilizing sound for communication is that crowd noise, just like at a real sporting event, can be so loud that the signal-to-noise ratio makes the audio signal difficult to detect. Since we use a filter-based approach to solve the problem, we can easily add-in other inputs to make the localization even more accurate and robust. Future research will explore combining the audio localization approach with the information sharing and visual feature-based approaches for solving the symmetric field problem.

6. Acknowledgments

The authors acknowledge the support of the U.S. National Science Foundation and Office of Naval Research for portion of this work.

7. References

[1] RoboCup Standard Platform League. Available: http://www.tzi.de/spl/bin/view/Website/WebHome. Accessed 2012 Sept 7.
[2] Aldebaran Robotics. Available http://www.aldebaran-robotics.com/en/. Accessed 2012 Sept 7.
[3] Quinlan M, Middleton R (2010) Multiple Model Kalman Filters: A Localization Technique for RoboCup Soccer. RoboCup 2009: Robot Soccer World Cup XIII, volume 5949 of Lecture Notes in Computer Science. pp. 276-287.
[4] Coltin B, Veloso M (2011) Multi-Observation Sensor Reseting Localization with Ambiguous Landmarks. AAAI Conference on Artificial Intelligence.
[5] Yang Y, Aitken V (2005) Uniform clustered particle filter for robot localization. American Control Conference. pp. 4607-4612.
[6] Malayeri A (2010) Robot Localization in Symmetric Environment. Master’s Thesis, University of Windsor, Canada.
[7] Förster T, Laue T, Müller J, Graf C, Böckmann A, Münder T (2012) B-Human Team Description for RoboCup 2012. RoboCup 2012: Robot Soccer World Cup XV Preproceedings.
[8] Howard A, Mataric M, Sukhatme G (2003) Putting the ‘I’ in ‘Team’: An Ego-Centric Approach to Cooperative Localization. IEEE International Conference on Robotics and Automation. pp. 868-874.
[9] Bader M, Brunner H, Hamböck T, Hofmann A (2012) Colour Histograms as Background Description: An Approach to overcoming the Uniform-Goal Problem within the SPL for the RoboCup WC 2012. Proceedings of the AUdtrian Robotics Workshop 2012 (ARW-12).
[10] Anderson P, Yusmanthia Y, Hengst B (2012) Natural Landmark Localisation for RoboCup. Proc. of the RoboCup Symposium.
[11] Murray J, Erwin H, Wermter S (2009) Robotic Sound-Source Localisation Architecture Using Cross-Correlation and Recurrent Neural Networks. Neural Networks. j. 22: 173–189.
[12] Mumulo E, Nolich M, Vercelli G (2003) Algorithms for Acoustic Localization Based on Microphone Array in Service Robotics. Robotics and Autonomous Systems. j. 42: 69–88.
[13] Hornstein J, Lopes M, Santos-Victor J, Lacerda F (2006) Sound Localization for Humanoid Robots - Building Audio-Motor Maps Based on the HRTF. IEEE International Conference on Intelligent Robots and Systems. pp. 1170-1176.
[14] Brindza J, Thomas A, Lee S, McDermid W, He Y (2012) UPennalizers Team Description for RoboCup 2012. RoboCup 2012: Robot Soccer World Cup XV Preproceedings.
[15] Lyon D (2010) The Discrete Fourier Transform, part 6: Cross-correlation. Journal of Object Technology. pp. 17-22.
[16] Cooley J, Tukey J (1965) An Algorithm for the Machine Calculation of Complex Fourier Series. Mathematics of Computation. j. 19: 297-301.
[17] Reid G, Milios E (2003) Active Stereo Sound Localization. The Journal of the Acoustical Society of America. j. 113: 113–185.
[18] Nakadai K, Okuno H, Kitano H (2001) Epipolar Geometry Based Sound Localization and Extraction for Humanoid Audition. IEEE International Conference on Intelligent Robots and Systems. pp. 1395-1401.
[19] Schenker L (1960) Pushbutton Calling with a Two-Group Voice-Frequency Code. The Bell System Technical Journal. j. 39: 235-255.