# 2012 CANON EXTREME IMAGING COMPETITION

## Visual awareness for a humanoid soccer robot

Peter Anderson (4th year Bachelor of Computer Engineering) and
Sean Harris (1st year PhD in Computer Science), ID: xxxx
The University of New South Wales, School of Computer Science and Engineering
Supervisor: Dr Bernhard Hengst

**Abstract:** Imaging technology plays an enormously important role in real-time robotics. Using sophisticated vision algorithms, these autonomous humanoid soccer robots are able to precisely identify their surroundings using extremely limited computing resources. In this project, a number of new student-led innovations helped catapult an Australian university to the top level of international RoboCup soccer.

**Comparison to Leaders in the Field / Student Contribution:** The RoboCup Standard Platform League attracts universities and research groups from all around the world, with 25 teams qualifying for the 2012 World Cup. The UNSW team (rUNSWift) placed third overall, and scored more goals during the competition than any other team. The rUNSWift RoboCup entry and the innovations described in this report are entirely student work.

**Related Publications:**

1. **Peter Anderson**, Yongki Yusmanthia, Bernhard Hengst, and Arcot Sowmya. Robot Localisation Using Natural Landmarks. In *Proceedings of the RoboCup International Symposium 2012*, Mexico City, Mexico, 18-24 June, 2012. (Nominated for best paper).
2. **Sean Harris, Peter Anderson**, Belinda Teh, Youssef Hunter, Roger Liu, Bernhard Hengst, Ritwik Roy, Sam Li, Carl Chatfield. Robocup Standard Platform League - rUNSWift 2012 Innovations. Accepted into *Australasian Conference on Robotics and Automation (ACRA) 2012*.
3. **Peter Anderson** and Bernhard Hengst. Fast Visual Odometry for a Humanoid Robot. Submitted to the *IEEE International Conference on Robotics and Automation (ICRA) 2013*.
4. **Peter Anderson**, Youssef Hunter and Bernhard Hengst. An ICP Inspired Unified Sensor Model with Unknown Data Association. Submitted to the *IEEE International Conference on Robotics and Automation (ICRA) 2013*.

# 1 Introduction

The RoboCup Standard Platform League (SPL) is an international robot soccer competition in which all teams compete with identical humanoid robots (the Aldebaran Nao). Each game runs for two 10 minute halves where two teams, comprising four Naos each, compete on a 6 x 4 metre soccer field. The robots perceive their environment primarily through their two forward-facing 30 fps cameras, under challenging real world conditions. There are no electronic or visual beacons used to aid the robots' situational awareness, apart from generic soccer field line markings. In 2012, for the first time, the field was symmetric about the half-way line with both sets of goal posts coloured yellow, as shown in Figure 1. All image processing is performed on-board the robots' low-power Intel Atom 1.6GHz processor, which was upgraded this year from a 500MHz Geode. The same processor simultaneously performs the robots' autonomous localisation, motion planning and behaviour functions, meaning that computational resources are extremely limited and fast image processing is critical for success.

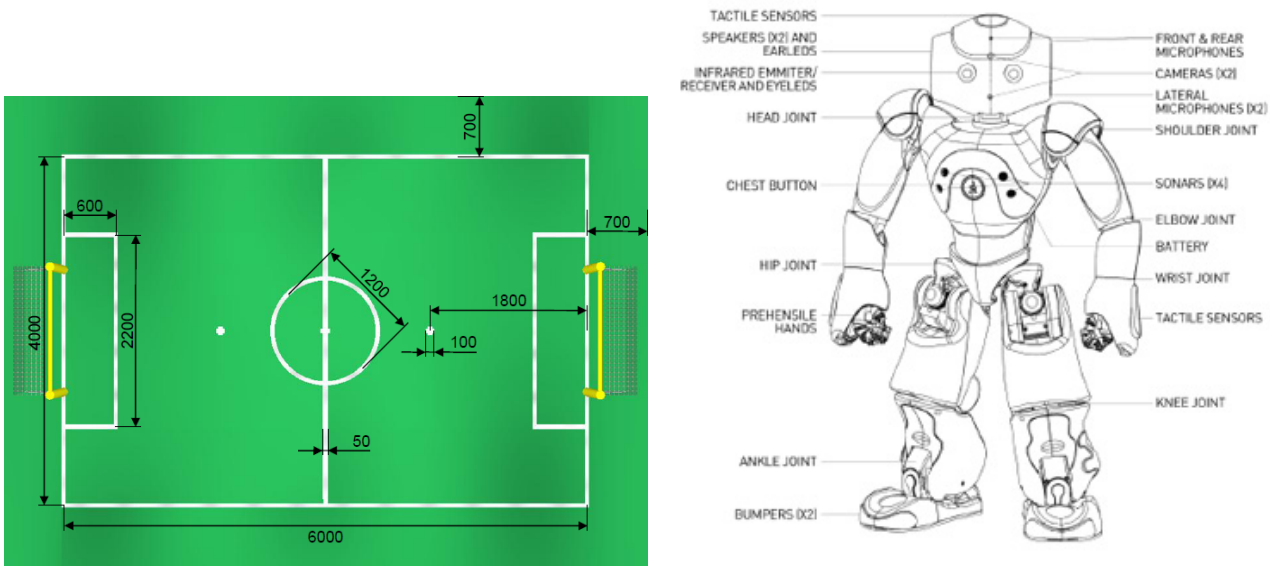



**Fig. 1.** Left: The RoboCup SPL field (dimensions in mm). Right: An Aldebaran Nao used in the competition.

Given the robots' severe resource constraints, it is not surprising that perception of the field environment remains one of the biggest hurdles facing many RoboCup teams. This project was motivated by the desire to achieve a level of visual awareness sufficient for each robot to be precisely localised at all times during the game, without needing to undertake specific localisation behaviours (like stopping to look around). After all, it is only after this objective

is attained that higher level strategic soccer behaviours, such as team formations, passing, and marking, will become most effective.

In pursuit of this aim, this paper outlines a number of new techniques that were developed in 2012 by the authors to improve the robots' perception of their environment. These enhancements include:

1. Exploitation of increased camera resolution and field of view,
2. Development of an extremely fast 1D SURF image feature,
3. A visual odometry module, and
4. A natural landmark localisation system,

All of these techniques were demonstrated in the 2012 RoboCup SPL competition, where the UNSW team (rUNSWift) placed third. Videos of rUNSWift's 2012 RoboCup matches can be found on the UNSW Computing Youtube channel[1].

## 2 Exploitation of Camera Resolution and Field Of View

Along with the upgraded processor, the Nao v4 introduced in 2012 offered improved camera resolution and allowed both cameras to be accessed simultaneously, effectively increasing the robots' vertical field of view from 34 degrees to 86 degrees. These hardware improvements offered a huge amount of potential to improve all aspects of the rUNSWift vision system. Access to both cameras at the same time allowed simultaneous tracking of the ball and field features. This in turn removed the need for any pause in game play to actively localise, since the robot is constantly receiving confirmation of its position. Figure 2 Left shows the extra features on offer when lining up the ball with two cameras instead of one.

In addition to this, the increased processing power allowed the resolution of object detection scans in the top camera image to be doubled. As a result of this, the robot is able to detect field lines up to 3m away, instead of only 1.5m away in 2011. It is also able to reliably track the ball 6m (length of the field) away instead of 4.5m in 2011. Figure 2 Right shows the goal keeper now being able to detect the centre circle from its own goal box, allowing it to stay localised without looking away from the ball. It also shows that the goal keeper can track the ball from the other side of the field, which helps the team maintain a strong belief about where the ball is at any point in time. A video showing detected field elements from the Nao's view during kidnap recovery tests is available [2].

[1] http://www.youtube.com/user/UNSWComputing
[2] https://dl.dropbox.com/u/36660950/ICPhiquality.mp4

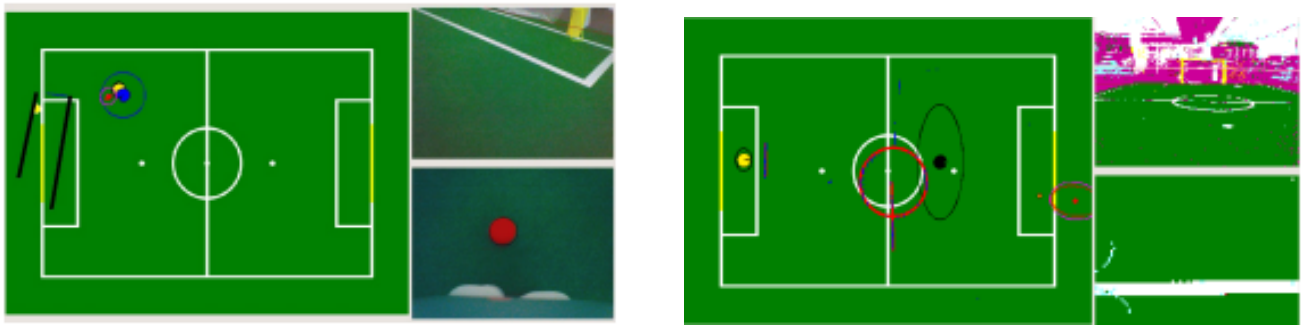


**Fig. 2.** Left: What the robot sees whilst shooting a goal. Right: What the goal keeper sees in a colour-classified image.

## 3 1D SURF Image Features

With the introduction of symmetric goal colours in 2012, a robot that becomes lost, for example after a complicated fall in the middle of the field, is not able to immediately resolve one end of the field from the other. This can result in robots changing teams midway through the match, with disasterous consequences! To overcome this problem, a 'kidnapped' robot needs to recognise landmarks in the unstructured environment around the field. This can be done by extracting scale-invariant local features from images, and finding feature correspondences, and ultimately a perspective transformation, between two images containing the same object.

SURF (Speeded Up Robust Features) [5], [4] and SIFT (Scale-Invariant Feature Transform) [9] are two existing methods for extracting invariant local features from images. However, these methods are relatively computationally expensive and difficult or impossible to implement in real time on a resource constrained robot. To overcome these resource limitations, an optimised feature detector consisting of a modified one dimensional variant of the SURF algorithm was developed (1D SURF). This method was then applied to a single row of grey-scale pixels captured at the robot's horizon, as shown in Figure 3. The horizon image was chosen for analysis because, for a robot moving on a planar surface, the identified features cannot rotate or move vertically. The use of a 1D horizon image and other optimisations dramatically reduces the computational expense of the algorithm, while exploiting the planar nature of the robot's movement and still providing acceptable repeatability of the features. To our knowledge only one group of researchers have published work that uses 1D image features for robot navigation [6], [7], [8], however they used SIFT features and benefitted from the availability of an omni-directional camera in their application.

Consistent with the original SURF algorithm, 1D SURF features are relatively robust to lighting changes, scale changes, small scene changes and small changes in viewing angle. As shown in Table 1, the recognition performance 1D SURF features extracted from the image


**Fig. 3.** Left: Image captured by the Nao robot showing superimposed 30 pixel horizon band in red, and the extracted grey-scale horizon pixels at the top of the image. Right: Matching 1D SURF features between two similar images, with outliers discarded using RANSAC.

horizon (evaluated based on the area under the ROC curve) is not as strong as SURF features extracted and matched across the entire image. However, 1D SURF executed more than one thousand times faster than SURF in this experiment. The mean execution time on the Nao v4 robot is 2 ms, which means the technique can be used in real-time on this hardware. Furthermore, as highlighted in Figure 4, the repeatability of 1D SURF features is still sufficient to largely resolve the SPL field-end ambiguity using just two stored images, at least in a relatively distinct environment with few scene changes. As such, 1D SURF features strike an excellent balance between repeatability and extreme speed.

**Table 1.** Running time and recognition performance of feature extraction algorithms on a 2.4GHz Core 2 Duo laptop.

| Feature extraction technique | Feature matching technique | Mean no. features | Mean extraction time (ms) | Mean matching time (ms) | Area under ROC curve |
|---|---|---|---|---|---|
| SURF | Nearest neighbour (NN) | 429 | 222.3 | 19.1 | 98.8% |
| 1D SURF | Nearest neighbour (NN) | 59.2 | 0.158 | 0.069 | 88.0% |
| 1D SURF | NN with RANSAC | 59.2 | 0.158 | 0.076 | 89.6% |

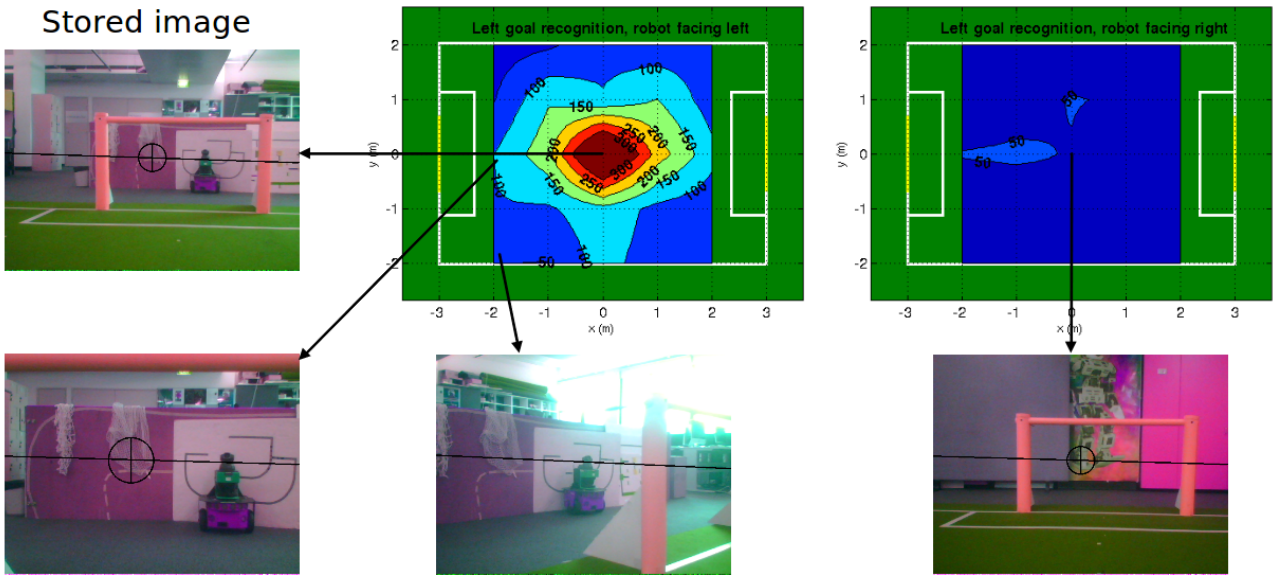



**Fig. 4.** Heatmap showing the recognition response in different areas of the field to 1D SURF features extracted from a single stored image.

Simultaneously with our own developments, other researchers have made significant progress developing fast binary image descriptors such as FREAK [1]. In future work, these fast binary descriptors might also be adapted to a 1D representation, potentially providing even greater efficiency improvements for mobile robots operating in planar environments.

## 4 Natural Landmark Localisation

The purpose of the natural landmark localisation system is to disambiguate one end of the soccer field from the other, to ensure that rUNSWift robots will never end up playing for the opposition team. Although 1D SURF features exhibit excellent recognition performance using simple nearest neighbours matching to a single image, as shown in Section 3, in practise the off-field environment consists mostly of spectators who tend to move around. As such, it is necessary to store many images of each goal and more scalable feature matching techniques are required.

With this in mind, a 'bag of words' natural landmark localisation system was developed that enables each robot to record a database of up to 40 images of each goal area while walking on to the field at the start of each half. A bag of visual words is a sparse frequency vector. It counts the number of occurances in that image of each word from a vocabulary of visual words pre-learned from typical features [10]. The retrieval system, outlined in Figure 5, is fast

enough to enable every camera frame to be checked against the database whenever a goal post is visible, with a mean retrieval time on the Nao v4 of 10.2 ms.

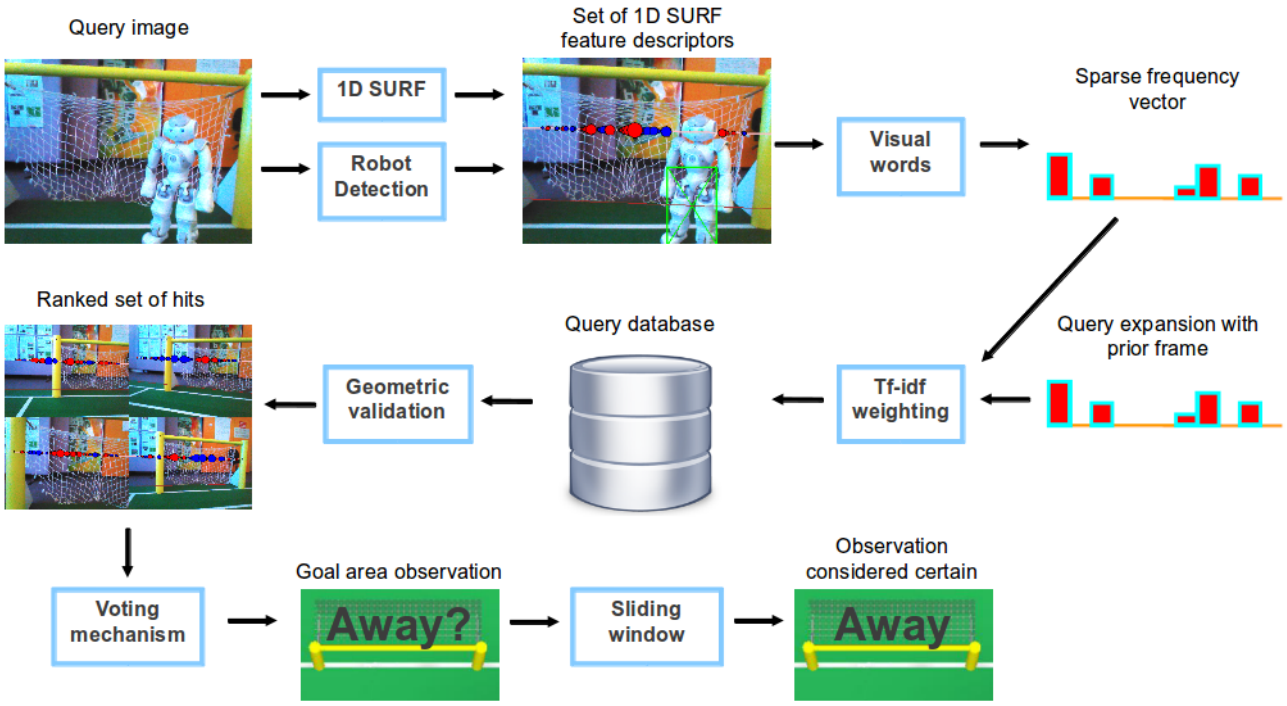
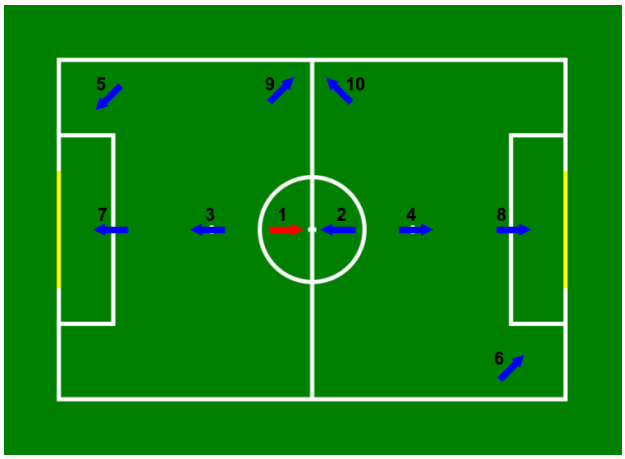



**Fig. 5.** Overview of the natural landmark localisation pipeline. At the end of the process, the robot can identifiy if it is facing it's home goal or the opposition (away) goal area, based on the landmarks in the background.

Key elements of the system include:

1. Subset-furthest first K-means cluster initalisation during visual vocabularly learning, which we found to produce better quality visual words,
2. The use of a stop list, to repress visual words that are too common and don't contribute to retrieval performance,
3. Query expansion, which significantly improves the performance of bag of words scene retrieval, at little cost [2],
4. Term frequency - inverse document frequency (tf-idf) weighting, which increases the weighting on visual words that appear frequently in the image, while decreasing the weighting on words that appear frequently throughout the whole database and are therefore not distinctive [3],
5. Geometric validation of short-listed image matches using RANSAC, and
6. A voting mechanism and a sliding window to filter observations over time.




| Trial | Time taken (s) | Correct field end (Y/N) | Positioning error (mm) |
|---|---|---|---|
| 1 | 0 | Y | 15 |
| 2 | 12.1 | Y | 65 |
| 3 | 12.9 | Y | 53 |
| 4 | 17.1 | Y | 25 |
| 5 | 32.1 | Y | 22 |
| 6 | 21.7 | Y | 95 |
| 7 | 16.9 | Y | 7 |
| 8 | 25.4 | Y | 28 |
| 9 | 16.6 | Y | 88 |
| 10 | 19.7 | Y | 42 |

**Fig. 6.** Left: Illustration of the positions used in the kidnap experiment. In each case the robot was required to return to the kick-off position labelled as position 1. Right: Kidnap performance of the natural landmark localisation system in repeated trials.

In order to evaluate the effectiveness of the complete natural landmark localisation system, a demonstration was devised which was also used as the rUNSWift SPL Open Challenge entry, which was awarded second place. During this demonstration, the robot is allowed 45 seconds to walk on to the field to the kick-off position, mapping landmarks as it walks. SPL matches also begin in this fashion. After the 45 seconds expires, the robot can be kidnapped to any field location and it is expected to return to the kick-off position on the correct side of the field. The robot is deemed to have arrived when it first stops walking. As illustrated in Figure 6, the robot was able to disambiguate each end of the soccer field flawlessly during 10 kidnap trials conducted in the rUNSWift lab. A video illustrating one of these kidnap trials is available[3].

## 5 Visual Odometry

Odometry refers to the ability of a robot to estimate its motion. It is used in the process update of the localisation filter to account for the robot's transition to a new position. In previous years rUNSWift odometry has been quite crude; effectively taking information from the walk-engine and assuming that the robot will move as directed. However, practical experience indicates that bipedal robots slip to varying degrees while walking, and often do not walk in straight lines! On the soccer field this is exacerbated as robots are often bumped by other robots and the goal posts, or impeded by the touching of arms or feet.

During these manoeuvres it is apparent that the greatest odometry inaccuracy is observed in the robot's heading, which can change very quickly, rather than in the forward or sideways

[3] https://dl.dropbox.com/u/36660950/VIDEO0002.mp4

component of the robot's movement. As the Nao is not fitted with a z-axis gyroscope, a visual method is required to correct inaccuracies in walk-engine heading odometry, and as always in this application domain, algorithms are subject to enormous resource constraints. With this motivation, a visual odometry module was developed using the same 1D SURF features already used by the natural landmark localisation system. By matching these features across subsequent frames, this module can accurately measure changes in the robot's heading over time.

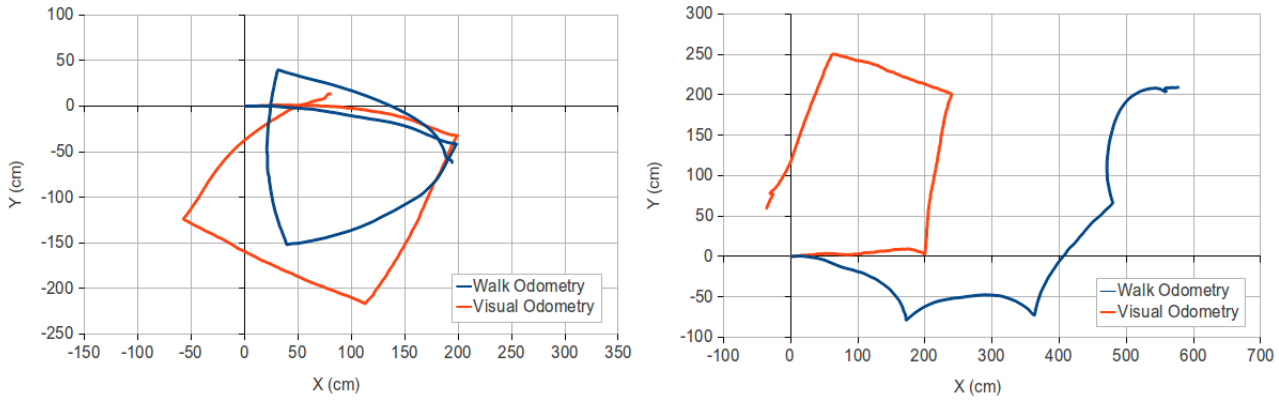



**Fig. 7.** Odometry tracks for one trial of an uncalibrated Nao robot walking in a 2 m x 2 m square path in both clockwise (left) and counter-clockwise (right) directions.

As illustrated in Figure 7, use of the visual odometry system results in a significant improvement in localisation accuracy. Repeated experiments indicate that the visual odometry system can reduce the odometric uncertainty of an uncalibrated Nao by 73%, while remaining robust to the presence of independently moving objects (such as other robots, or the referee). Furthermore, by analysing heading discrepancies between walk-engine odometry and visual odometry, obstacle collisions on the left and right sides of the body can be immediately detected and the appropriate avoidance behaviour taken. For example, it was observed that allowing the robot's arms to become limp during a collision avoided a significant number of falls at competition, as the robot was more easily able to brush past other robots.

Extreme efficiency is achieved by exploiting the planar motion assumption in both the feature extraction process (using 1D SURF) and in the pose estimation problem, which seeks to find the heading change between frames by analysing the displacement histogram of feature matches, resulting in a mean execution time on the Nao v4 of 4.5 ms. The visual odometry system makes no assumptions about the nature of the environment other than planar motion, and so it is potentially applicable to a wide range of mobile robot navigation problems. A video

illustrating the robustness of the visual odometry system in the presence of an independently moving objects is available[4].

## 6 Conclusion

We have presented a number of related innovations in imaging for real-time robotics, with results that indicate these systems operate extremely efficiently and effectively. However, the true performance test for a large scale integrated system is running it under challenging real world conditions, seeing if it achieves its objectives, and observing how it compares to other competitors. In this case that means on the soccer field under the pressure of competition matches.

In 2012 rUNSWift finished in 3rd place overall and 2nd in the Open Challenge, a big improvement over the 2011 results where the team was eliminated in the quarter finals. rUNSWift was also the top goal scoring team in the SPL, scoring a total of 62 goals over 8 games. This was again a huge improvement on the 2011 results in which rUNSWift scored 23 goals over 6 games. The team also did not score any own goals, which was a big achievement for the imaging system considering this was the first year in which the goal posts at each end of the field were the same colour. Anecdotal evidence from watching many games suggests that for the vast majority of game time, rUNSWift robots were well-localised. Furthermore, this was achieved without any of the localisation behaviour 'crutches' that were previously used, such as head-scanning before kicking the ball.

Figure 8 shows the results of the semi finals onwards, with the top 4 teams coming from the University of Texas at Austin (Austin Villa), the University of Bremen (BHuman), the University of Leipzig (Nao-Team HTWK) and the University of New South Wales (rUNSWift).

[4] https://dl.dropbox.com/u/36660950/VOhiquality.mp4

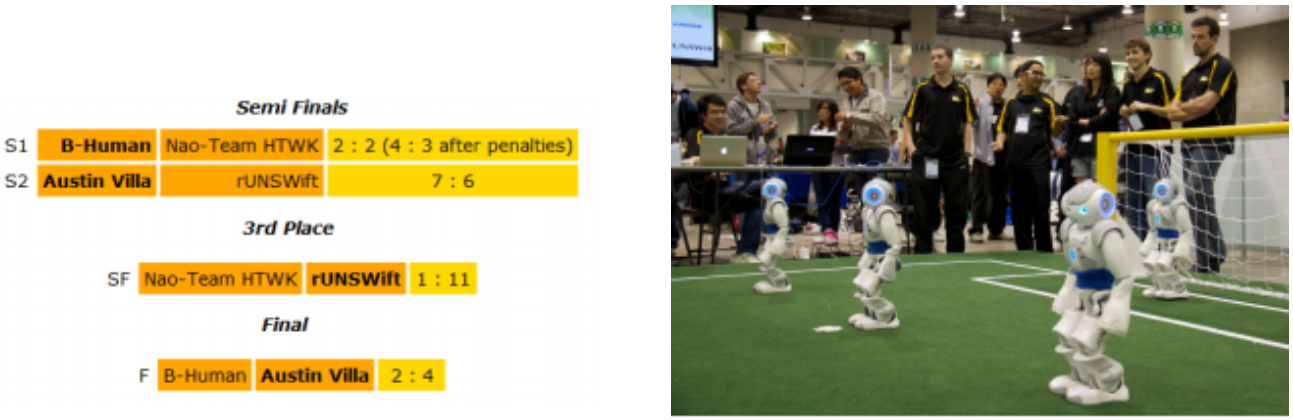



**Fig. 8.** Left: RoboCup Standard Platform League 2012 results. Right: rUNSWift ready for action in RoboCup 2012.

# References


1. Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak: Fast retina keypoint. *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, June 16-21, 2012*, pages 510–517, 2012.
2. R. Arandjelović and A. Zisserman. Three things everyone should know to improve object retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
3. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. ACM Press, 1999.
4. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.
5. H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404–417, 2006.
6. A. Briggs, C. Detweiler, P. Mullen, and D. Scharstein. Scale-space features in 1d omnidirectional images. In *Omnivis 2004, the Fifth Workshop on Omnidirectional Vision, Prague, Czech Republic*, pages 115–126, 2004.
7. A. Briggs, Y. Li, D. Scharstein, and M. Wilder. Robot navigation using 1d panoramic images. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 2679–2685. IEEE, 2006.
8. A.J. Briggs, C. Detweiler, Y. Li, P.C. Mullen, and D. Scharstein. Matching scale-space features in 1d panoramas. *Computer vision and image understanding*, 103(3):184–195, 2006.
9. D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
10. Josef Sivic and Andrew Zisserman. Video google: A text retrieval approach to object matching in videos. pages 1470–1477, Nice, France, 2003.