

The University of New South Wales
School of Computer Science & Engineering



Natural Landmarks Localisation

Yiming Deng (3290611)
Supervisor: Dr. Bernhard Hengst

August 2011

This report is submitted in fulfillment of the requirement of
Special Project A (COMP3901)

Abstract

This report considers two techniques in determining the probability distribution of a robot's heading with respect to a predefined map. These techniques examined are called Region binning and Cross-Correlation and will be employed in the context of Robot Soccer. Sample images of the horizon will be extracted from a robot standing on a standard Robocup soccer field in order to determine the corner of the field it is facing.

Region Binning and Cross-Correlation have the potential to develop into something practical. However these techniques are not robust when dealing with distortion and other various field-related errors. Addressing these issues is vital for these techniques to be of practical use.

Contents

1	Introduction	3
1.1	Goals	3
1.2	Contribution	3
1.3	Report Structure	3
2	Background	4
3	Theory	6
3.1	Preliminary & Definitions	6
3.2	Region Binning	6
3.3	Cross-Correlation	7
4	Experiment & Results	7
4.1	Experiment Method	7
4.2	Sample Images	8
4.3	Results	8
5	Analysis	11
5.1	Accuracy	11
5.2	Computational Performances	12
5.3	Errors	13
6	Conclusion	13
7	Bibliography	14
8	Appendix	15
8.1	Stereo Vision Technique	15
8.1.1	Method	15
8.1.2	Results	15
8.1.3	Analysis	16
8.2	Modified Edit Distance Algorithm	18

1 Introduction

In previous Robocup SPL competitions the game was always played on the field. Anything above the field was treated as noise and deemed too random by nature to be practically useful. In particular in the early days of Robocup there used to be beacons off the field to aid localisation. They eventually disappeared and attention then turned to localising off the different coloured goal-posts. Now that both goal-posts will be uniformly coloured from next year onwards there is a need to localise off other features of play. Our motivation therefore is to utilise some features above the field to improve the localisation of the Nao robot.

Although the methods and techniques presented in this report can be applied to quite general situations the environment of the experiments presented will be on a soccer field inside a research laboratory, the robot being the Aldebaran Robotics humanoid Nao.

1.1 Goals

The goal of this report is to provide, from any position of the soccer field, $P(z|x)$ - the probability distribution where z is the heading of the robot if it was standing on the center of the field and x is the 360-degree view of the horizon from the center.

1.2 Contribution

This report provides an analysis on two techniques, Region Binning and Cross-Correlation, and how they process information from the horizon to provide a probability distribution of the corner of the field a robot is facing.

1.3 Report Structure

The rest of the report is structured as follows: section 2 will provide a background of similar techniques in the area. Section 3 will provide the theory behind the Region Binning and Cross-Correlation techniques. Section 4 will provide details on the experiments performed along with their results which is followed by section 5, an analysis of the results. Finally the last section is the conclusion, addressing the techniques, goals and possibilities of future work in this area.

2 Background

On a large scale Natural Landmarks localisation is a subset of Simultaneous Localisation and Mapping (SLAM). In order to take advantage of landmarks surrounding the field some feature detection would be ideal. Lowe's SIFT algorithms[1] provides accurate feature detection by passing through a series of techniques which aid robust object detection. These techniques however are not ideal for the Naos as they are computationally too inefficient.

On a simpler level landmark localisation incorporates edge detection of objects, in which there is a wealth of literature dedicated to finding accurate yet computationally cheap algorithms. Marr and Poggio[3] outlined in their seminal paper a way of detecting edges in Greyscale images by defining an edge as having a local maximum/minimum at the first difference equation and a zero-crossing point in the second difference equation. Figure 1 demonstrates an illustration of this. The Sobel Operator is another measure of edge strength where a scalar for each pixel is calculated by passing a kernel filter through their image intensity values. For a 3x3 kernel a typical matrix is

$$\begin{pmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{pmatrix}.$$

This measures edge intensity in the horizontal direction. The vertical direction intensity would then be calculated using the below matrix:

$$\begin{pmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{pmatrix}.$$

Using the ideas of edge detection a stereo vision method of Natural Landmarks localisation has been devised. By hard coding initial position coordinates inside the robots, landmarks can be localised using stereo vision methods. Then a disoriented robot can be relocalised using triangulation methods. More details of this method are located in the appendix. However this method is not robust when detecting landmarks with similar colours and its landmark detection is sensitive to small measurement errors.

Therefore a new perspective of utilising natural landmarks is proposed - one which considers the relative order of landmark colours observed by the robot in order to perform localisation.

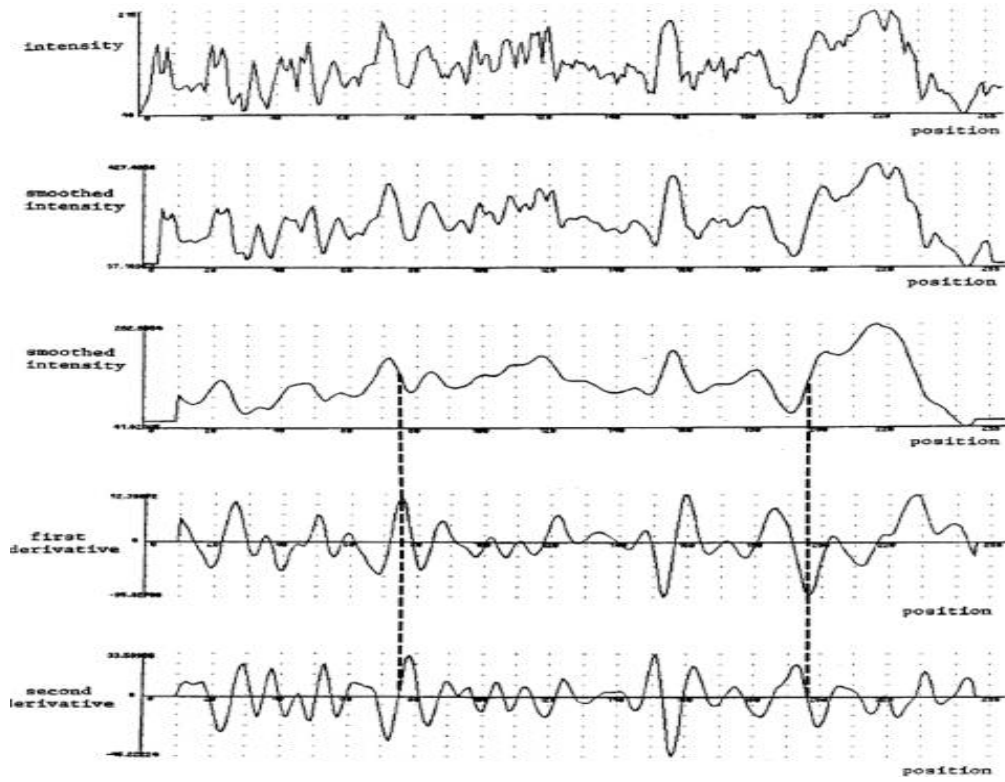


Figure 1: Edge detection using finite differencing by Marr and Poggio.

3 Theory

3.1 Preliminary & Definitions

In this section an overview of the theory behind the Region Binning and the Cross-Correlation technique is presented. For the rest of this report define the initial ring of pixels as the 360-degree view of the horizon from the centre of the map, represented by YUV pixels. Also define the current horizon pixels as the pixels on the horizon of the current frame on the robot camera. Define the centre-heading to be the heading of a robot if it was placed on the center of the field.

In both techniques the aim is that a disoriented robot on the soccer field can use the current horizon pixels to work out the center-heading using an algorithm, given the initial ring of pixels. The centre-heading is not the heading of the robot, but intuitively it is the corner of the map that a robot is facing. All angle measurements are measured by the centre-heading because it provides a measure that is relative to the initial ring of pixels and irrespective of its field position, so between different robot frames the corner of the field a robot is facing can be determined.

3.2 Region Binning

Region Binning consists of dividing the initial ring of pixels into a small number of regions. Mathematically this means the 360-degree arc in the initial ring of pixels is divided into n arcs of equal angle, where n is the number of regions. Each region represents a corner of the map, having a set of pixels associated with it. The distribution of the centre-heading is then decided by a voting process, described by the following pseudo-code algorithm

```
for every pixel in current_horizon_line
  for every region in regions
    for every pixel_region in region
      if equal(pixel, pixel_region)
        region->regionvotes++;
```

where `equal(a, b)` tests whether two pixels are equal based on thresholding the sum of the difference of each YUV pixel value to a predefined number. Experiments during stereo vision testing have shown that a threshold value

of 30 provides an optimal balance between detection of objects in most situations. The votes of the regions make up the probability distribution of the centre-heading.

3.3 Cross-Correlation

The idea of Cross-Correlation is to take the current horizon line and look at where it could be on the initial ring of pixels. To do this the current horizon line has to be compared to the initial ring of pixels at every displacement, and each comparison needs to be measured with a value. The pseudo-code is as follows:

```
int i, initPixNum = size(initial_ring_of_pixels);

for (i = 0; i < initPixNum; i++)
    for (j = 0; j < size(current_horizon_line); j++)
        cross_corr[i] +=
            diff(initial_ring_of_pixels[(i+j)%initPixNum],
                current_horizon_line[j]);
```

where `cross_corr` stores the cross-correlation function and `diff(A, B)` is the function to calculate the sum of the differences in YUV values between pixel A and B. Therefore the best match occurs at the global minimum of `cross_corr`. Inverting the order of the `cross_corr` values gives a measure of the distribution of the centre-heading.

4 Experiment & Results

4.1 Experiment Method

The experimental method is as follows:

- Set up the horizon line detection on the robot. For more information about the setup refer to the 2010 rUNSWift Report[4].
- Set up the initial ring of pixels. Place the Naos on the centre of the field. Define zero degrees as the direction of the left corner facing the opponent's goal. Starting from zero degrees, record nine different pixel arrays of the horizon line, each spanning 40 degrees. This should be

done by aligning the left border of the camera image at zero degrees, then recording one array, then turning the robot on the centre spot so that the left border faces 40 degrees and record again, and so forth. Then stitch the nine arrays together to form the initial pixel of arrays. Note that nine arrays of 40 degrees was chosen because a standing Nao's horizon range on one camera frame is slightly in excess of 40 degrees.

- Record sample images from the Nao facing in random directions, standing anywhere on the soccer field. For this particular purpose the images are recorded and stored via OffNao. Refer to the rUNSWift report[4] about OffNao.
- Run the Region Binning and Cross-Correlation Algorithms on the set of sample frames to produce the centre-heading distributions for analysis.

4.2 Sample Images

In order to demonstrate the performance of the algorithms in a suitable environment the results of the experiment was performed in the AI Research Labs in the Faculty of Computer Science and Engineering K17 building. In particular the sample images were generated from a standing Nao Robot on a standard Robocup soccer field, observing a surrounding environment of chairs, desks, boxes and cables. 200 sample frames of the surrounding environment based on random field positions and headings was collected. The emphasis of the sample images has been on extracting different angles and distortions of the same image of pattern colours for analysis.

4.3 Results

Figures 2 and 3 show four images representing different angles of roughly the same view. Below each image is the centre-heading distribution of the Region Binning and the Cross-Correlation technique. The red line in each distribution represents the ground truth centre-heading.

A summary of statistics based on the accuracy of the techniques performed on the 200 frames is given in the first table. Accuracy in the following is measured for each frame based on whether there is a local or global maximum in the centre-heading distribution. A ratio of the local to global

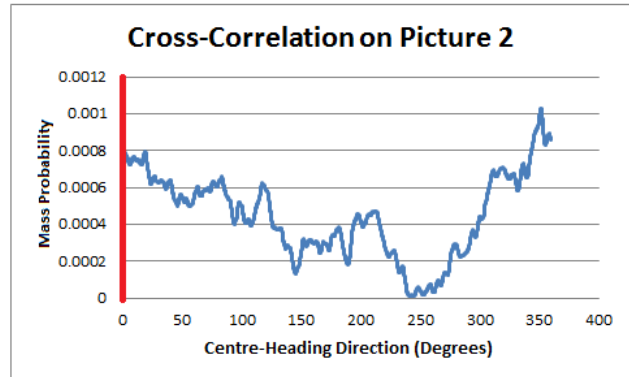
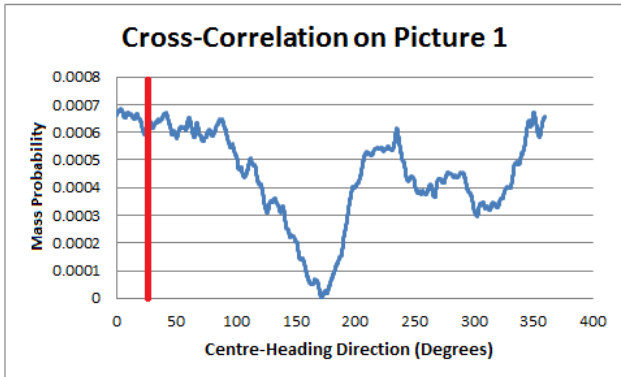
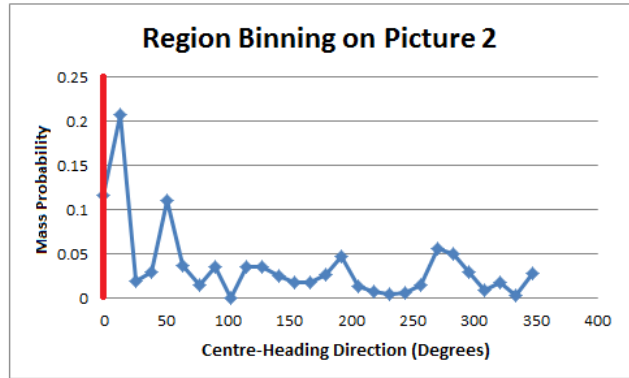
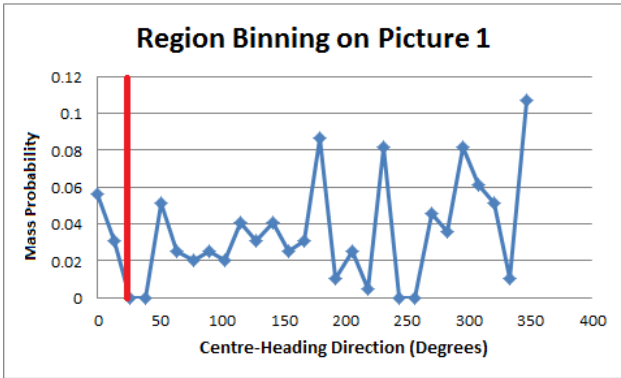
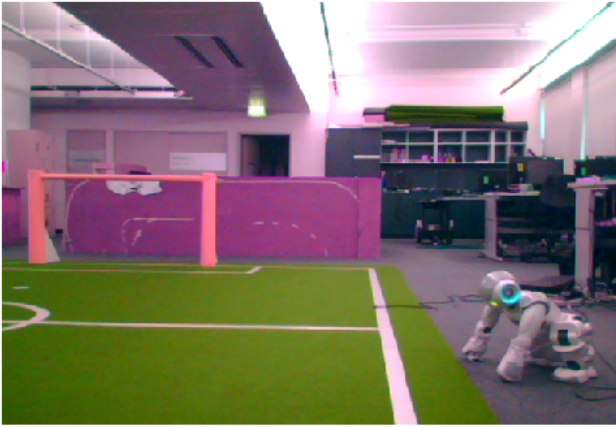


Figure 2: Pictures 1 and 2 with their distributions.

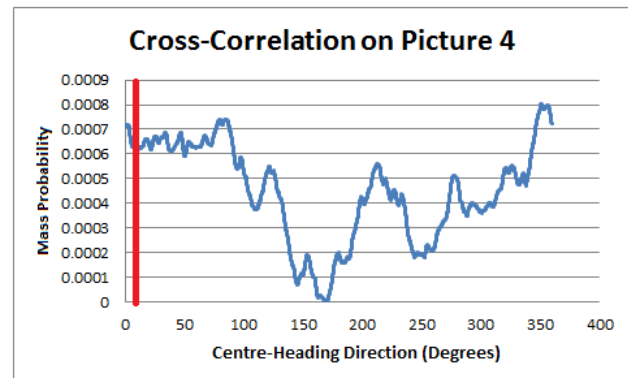
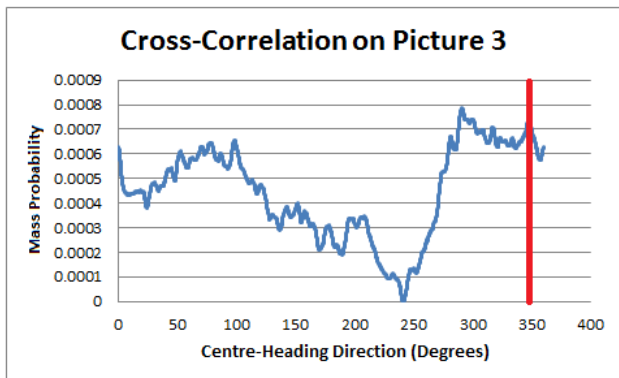
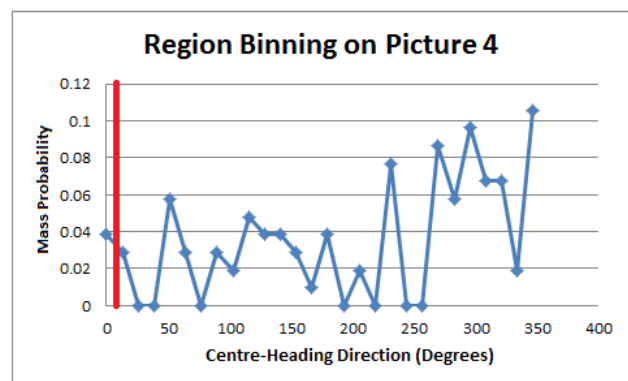
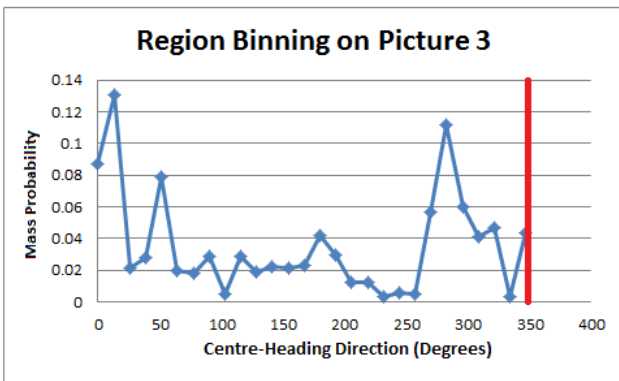


Figure 3: Pictures 3 and 4 with their distributions.

maximum percentages are included in the table. The ratio is an indicator of how prevalent a global maximum is when a local maximum is found. In other words, it calculates the sample probability of a global maximum, given a local maximum.

Method	Local Maximum (%)	Global Maximum (%)	Local/Global Ratio
Region Binning	70.50	22.00	31.21
Cross-correlation	82.50	45.50	55.15

The below table provides basic performance statistics of processing each frame with each algorithm and also with no algorithm, in order to test the performance of overheads.

Method	Average Speed (μs)	Sample Standard Deviation (μs)
Region Binning	8614.17	2437.99
Cross-correlation	23921.67	2534.13
None	40.63	21.74

5 Analysis

5.1 Accuracy

The implementation of the techniques on the four pictures in the results section was meant to test the algorithms on different camera angles of the same region, however the algorithms returned mixed results overall. Except for cross-correlation on picture 3, the techniques did not return maxima for any of the pictures. A localisation method is considered robust if the method can return similar results based on the detection of the same pattern of objects, and the algorithms are inadequate in this regard.

However it must be noted that in the Cross-Correlation distributions the ground truth is located in the middle of several consecutive high probability mass values. Because the Cross-Correlation function has a discrete unit measurement of one pixel, consecutive cross-correlation values are often similar. Even if the measurement is off by a small number of pixels, as long as the pixel colour ordering is preserved then consecutive cross-correlation values should be similar. This implies that it might be more accurate to find a

range of centre-headings to determine the ground truth. A flaw in this idea however is that these ranges might not be the most prominent clusters, as the results suggest. So finding a robust algorithm for this would also be a challenge.

For the Region Binning method, a susceptibility of the algorithm to distortion might be a major factor in its ability to accurately determine the ground truth. In picture 1, the right hand side of the frame represented a heavily distorted view of one side of the field, which would be quite different from the initial ring of pixels, taken from the centre of the field. This would mean that on one hand, the current horizon line would match many regions in the voting algorithm, because pixels of many regions are represented in the current line. On the other hand there would be few votes for all regions because there is not enough pixels belonging to one dominant region of the map, because of distortion. The Region Binning method would work best if the current horizon line had pixels belonging to one distinct region of the initial ring of pixels, but judging by the results, the algorithm lacks robustness.

Looking at the table of maxima detection there is a high percentage of maximum detection in both methods and less so for global maximum detection. The low local/global ratios mean that maxima detections were predominantly local and because of this a large number of false positives would arise if the ground truth was based on the global maximum. Of the two methods Cross-correlation has both a higher rate of maxima detection and a local to global ratio, meaning for the sample it yield more accurate probability distributions. However it must be noted that a large portion of the samples were on clear images which had distinct colour patterns in its surroundings. Restriction of the sample images to the lab environment hindered the reliability of the results in a practical scale.

5.2 Computational Performances

The average speed of the Cross-Correlation has be shown to be about thrice as fast as the Region Binning. This comparison is reasonable because of the extra processing the Cross-Correlation method needs to do to create a more refined probability distribution. The processing time for overheads seems insignificant compared to the processing time for the algorithms. Overall it seems that in order to get a more refined and accurate probability distribution there is a speed penalty involved.

5.3 Errors

There were several errors encountered with the experiment that affected the quality of our results. As previously mentioned the distortion of images seen from different positions led to different probability distributions. Also depending on the robot's position on a field, the order of objects between the different images could be seen differently. This would lead to a different ordering of colours than the one expected on the initial ring of pixels. Another practical error source could be field symmetry, where if large regions of the map are similar on the horizon then the algorithms would return multiple ground truth possibilities. In the worst case of a horizon line which is mostly uniform in colour the methods described would be invalid. Addressing these errors in future experiments would make the algorithms more robust to a practical environment.

6 Conclusion

The experiments performed reached its goal in providing the probability distribution of the centre-heading of a robot given the initial ring of pixels, based on the Region Binning and Cross-Correlation methods. Relatively, the Cross-Correlation method gave a more refined probability distribution over the Region Binning method at a slower speed and was also shown to be more robust. Both methods show promise that they can provide some information of the direction of the field a robot is facing regardless of its position on the field.

However due to a variety of distortion and other field-related errors these techniques are far from practical use. The algorithms need to be more robust, so that no false positives arise in their results. Other than directly modifying the algorithms to address these errors, other ideas include finding a heuristic to deal with the region of detected colours in the binning methods and to look at cheaper forms of scale-invariant edge detection techniques. Another improvement that can be made is to look not only at the horizon pixels but at a larger subset of the camera image, because in general more information can be extracted if the given dataset is larger.

7 Bibliography

- [1] D. Lowe, *Object Recognition from Local Scale-Invariant Features*. The Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 2 (06 August 1999), pp. 1150-1157.
- [2] S. Lazebnik, C. Schmid and J. Ponce, *Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories*. In CVPR, Vol. 2 (2006), pp. 2169-2178.
- [3] D. Marr and T. Poggio, *A computational theory of human stereo vision*. Proceedings of the Royal Society of London. Series B, Biological Sciences, Vol. 204, No. 1156. (1979), pp. 301-328.
- [4] Jayen Ashar, David Claridge, Brad Hall, Bernhard Hengst, Hung Nguyen, Maurice Pagnucco, Adrian Ratter, Stuart Robinson, Claude Sammut, Benjamin Vance, Brock White, Yanjin Zhu, *RoboCup Standard Platform League - rUNSWift 2010*. Australasian Conference on Robotics and Automation, Dec 1st-3rd Brisbane Australia, 2010

8 Appendix

8.1 Stereo Vision Technique

8.1.1 Method

1. Two robots are placed at known positions with known headings, facing one image.
2. The pixels on the horizon will be recorded in its YUV form.
3. Edge detection will be done. Between consecutive pixels the absolute difference between their Y, U and V values is recorded. An edge is recorded if their absolute difference is greater than a threshold number. Empirical testing shows $\Delta(Y) + \Delta(U) + \Delta(V) \geq 30$ is a good threshold condition.
4. Edge filtering is done to filter out noisy and thick edges.
5. From the edges colour bands can be extracted for both robots.
6. The two colour bands are then passed through the modified edit distance algorithm (see later on in the Appendix) to find a subset of matching colour bands that both robots can see. These colour bands are the landmarks that both robots have detected.
7. From these colour bands, the robots perform triangulation to localise each landmark on the map.
8. Whenever one of these two robots is moved to a new position. If it detects these landmarks again it can use these landmarks to find possible positions of its current location on the field map.
9. These clusters can be grouped to form a possible robot location on the map.

8.1.2 Results

Edge detection using the gradient method and threshold was reliable, particularly in images where the colour transition was clear. On small images further away it was also reliable, but became less so when the images were

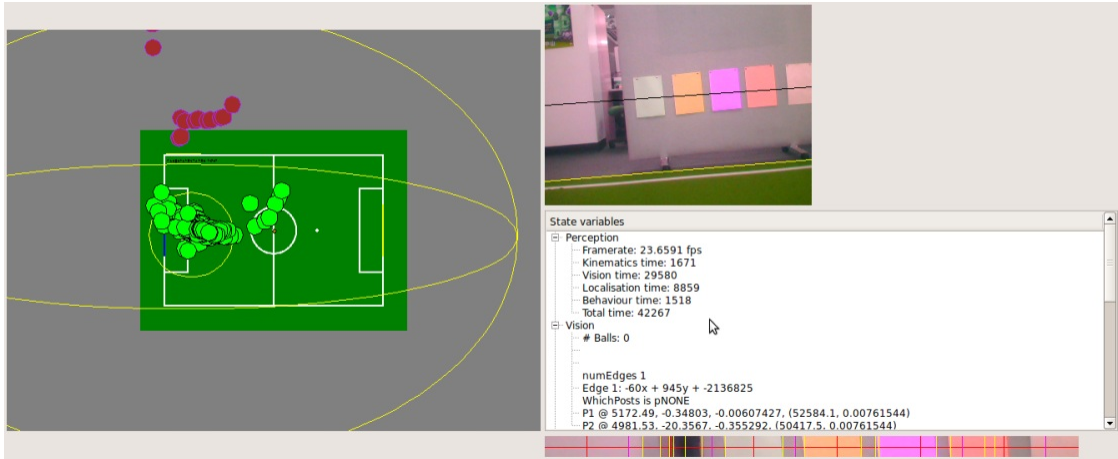


Figure 4: Results from localising off coloured paper on a billboard.

small (for example wires and cables). Also when the robot was moving there was no detection at all. This might be because of the blurry image which was observed. The fast movement might make YUV values for all consecutive pixels very similar.

Landmark localisation initially was accurate for objects with distinct separate colour regions whilst it was less accurate for similar colours. Landmarks localisation relied mainly on edge detection and thus shares most of its advantages and disadvantages with it also. Furthermore detecting many natural landmarks like poles, pillars, desks, chairs and boxes far away resulted in wildly inaccurate measurements.

The many landmarks detected from the above paragraph meant that there were many possible robot coordinates calculated. The number of robot coordinates detected varied greatly, from none to 400. In most results a dominant cluster could be found, in other cases most results lie on an arc of the circle which intersected a majority of landmarks. Yet some cases still existed where a dominant cluster could not be found.

8.1.3 Analysis

Before the gradient method was implemented both methods in the background were tested. Marr and Poggio's method of differencing the respective YUV values only gave very few results. Upon looking at the YUV values

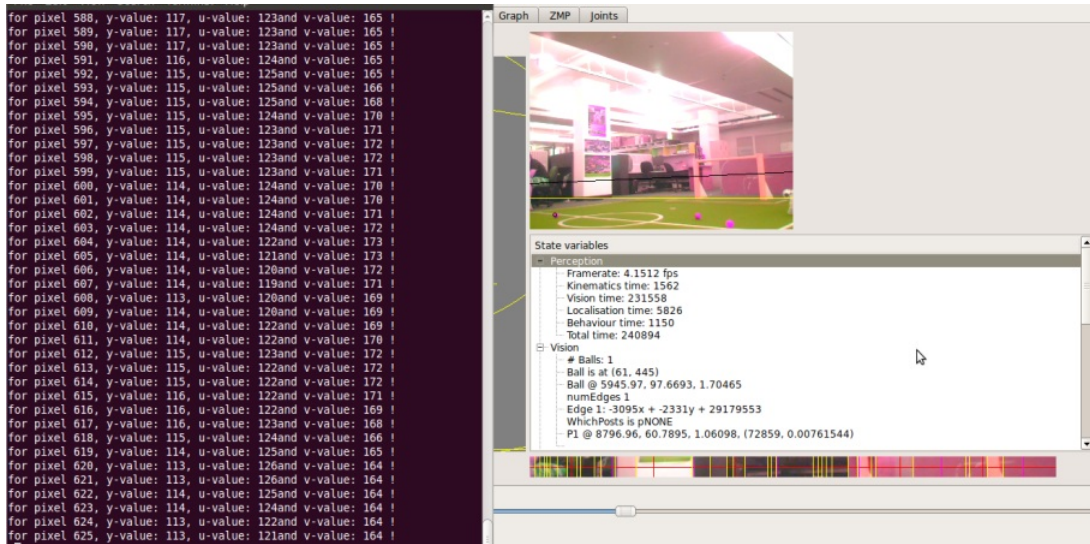


Figure 5: Some YUV values of consecutive pixels shown with a sample image.

this was probably because the changing YUV values were not enough to be picked up by the first and second difference equations. It could not detect clear edges. Noisy pixels were also significant enough for it to random detect edges which were clearly not there.

The Sobel operator did a much better job, but compared to the gradient method, it detected lesser edges. It was found that the Sobel/Scharr Operator was less flexible in its adjustments to edges and could not detect all of them, whereas the gradient method could be changed by changing its threshold value.

As mentioned before landmark localisation worked well for distinct ('unnatural') landmarks but did not work well in the practical scenario. One possible cause is that from the image several colour bands are similar, especially brown boxes and white poles and pillars. This could in turn confuse the modified edit distance algorithm into calculating an incorrect result as some colour band sequences are particularly similar, but not equal. Other prominent factors include measurement error, whether that comes from the pose that calculates the horizon, or error from the initial hard-coding of the primary robot positions and headings from which the landmarks can then be relatively detected. But perhaps the most significant error is error from measuring the subtended angle of the landmark from the image, as small

error in the angle would create huge error in landmark location.

Robot localisation from natural landmarks relied on both landmark detection to be accurate, which in turn relied on edge detection to be accurate. So if there was huge error in one of the two, the error would be compounded in robot localisation. When there was good landmark localisation many clusters could be found, and filtering techniques were often accurate enough to decide a single point. When landmarks were not accurate, possible robot locations were often dispersed around the map, a result of the compounding error.

8.2 Modified Edit Distance Algorithm

Problem: Let $0 \leq m \leq \max(k, l)$ and S be the set of all ordered colour bands. Given a sequence of two colour bands $A = a_1a_2a_3\dots a_k$ and $B = b_1b_2b_3\dots b_l$, $A, B \in S$, we wish to find $C = c_1c_2c_3\dots c_m \in S$, the sequence of colour bands which will require the least deletions for a_i 's and b_j 's from A and B combined.

So given A and B above we define the an edit distance matrix E , a $(k+1)$ by $(l+1)$ matrix. For the $(i, j)^{th}$ element $e[i, j]$, we further define $\delta(e[i, j])$ to be the minimal edit distance of colour bands $a_1a_2a_3\dots a_{i-1}$ and $b_1b_2b_3\dots b_{j-1}$ and $s(e[i, j])$ to be the colour bands that determine $\delta(e[i, j])$.

The initial conditions for matrix E are that $\delta(e[0, 0]) = 0$, $s(e[i, 0]) = \emptyset$ and for all $i \in (1, 2, \dots, k+1)$, $\delta(e[i, 0]) = i$, $s(e[i, 0]) = a_1a_2\dots a_{i-1}$ and $j \in (1, 2, \dots, l+1)$, $\delta(e[0, j]) = j$, $s(e[0, j]) = b_1b_2\dots b_{j-1}$. Define further $\beta(e[i, j])$ to be the largest consecutive set of recent deleted elements $b_jb_{j-1}\dots b_{j-c}$ in determining $s(e[i, j])$ and $\alpha(e[i, j])$ to be the same $a_ja_{j-1}\dots a_{j-d}$ in determining $s(e[i, j])$. $s(e[i, j])$ and $\delta(e[i, j])$ are determined as follows:

```

if (a_i in alpha(e[i-1, j-1]) and b_j in beta(e[i-1, j-1])) {
    Dist_1 = delta(e[i-1, j-1]);
} else {
    Dist_1 = delta(e[i-1, j-1])+2;
}

```

```

if(a_i in alpha(e[i-1, j]) ) {
    Dist_2 = delta(e[i-1, j])-1;
} else {
    Dist_2 = delta(e[i-1, j])+1;
}
if(b_j in beta(e[i, j-1])) {
    Dist_3 = delta(e[i, j-1])-1;
} else {
    Dist_3 = delta(e[i, j-1])+1;
}

delta(e[i,j]) = min(Dist_1, Dist_2, Dist_3);

if (delta(e[i,j]) == Dist_1) {
    if(Dist_1 == delta(e[i-1, j-1])) {
        s(e[i,j]) = s(e[i-1,j-1]);
    } else {
        /* without a_i and b_j */
        s(e[i,j]) = s(e[i-1,j-1]);
    }
} else if(delta(e[i,j]) == Dist_2) {
    if(Dist_2 == delta(e[i-1, j])-1) {
        s(e[i,j]) = s(e[i-1,j]);
    } else {
        /* without a_i */
        s(e[i,j]) = s(e[i-1,j]) ;
    }
} else {
    if(Dist_3 == delta(e[i, j-1])-1) {
        s(e[i,j]) = s(e[i,j-1]);
    } else {
        /* without b_j */
        s(e[i,j]) = s(e[i,j-1]);
    }
}
}

```

From here the sequence of colour bands that solves the problem is in $s(e[k + 1, l + 1])$ with edit distance $\delta(e[k + 1, l + 1])$.