# Natural Landmarks Localisation

Yiming Deng

February 13, 2011

**Abstract**

In this report the idea of natural landmark localisation is studied in some detail. In particular localising using two robots using triangulation techniques on detected landmarks on the horizon is examined. Furthermore profiling techniques based on comparing the current horizon to a predetermined 360-degree array representing the full horizon from a central point is observed.

Whilst triangulation is theoretically robust and has the most potential, it is not practical because of the amount of image noise and movement in a live game that can throw it off. Profiling has more potential to be practically useful, but more analysis and implementation must be made to make it game-ready.

# Contents

# 1 Introduction

In previous Robocup SPL competitions the league was always played on the field, that is, based on objects that we can recognise such as other robots, goal-posts and field lines. Anything above the field was treated as noise and deemed too random by nature to be of any use to the game. Our motivation thus, is to utilise some features, or natural landmarks, above the field to improve the information space of the Nao robot and its surroundings.

During this Taste of Research (ToR) experience, two techniques have been examined. The first technique, which from now on will be called the Stereo Vision technique, is by replicating stereo vision by using two robots to observe a picture from two different positions. From there possible localisation points can be retrieved. The other technique, which will be called the Profiling technique, implemented is by recording the 360-degree view of the horizon at one point, robots can determine the direction they are facing by comparing their current horizon vision with this predetermined ring of pixels.

In the consequent sections of this report some ideas related to techniques used will be given in the Background section. This will be followed by the method and its results, along with a discussion and a conclusion of the findings, which will detail suggestions on further improvements of the two techniques given.

# 2 Background

One of the key ideas towards Natural Landmarks Localisation is the property that pixels of objects detected on the horizon do not move vertically. Thus regardless of the robot's position on a map, if the same object is detected on the horizon, that same horizon pixel will be found on the robot's vision horizon line. The horizon line is generated from the robot's pose and will be prone to measurement and kinematic errors. But these errors are small enough so as to still allow us to make valid measurements based on the horizon property stated above.

In order to localise using stereo vision, edge detection needs to be performed. Marr and Poggio (1977) outlined in their seminal paper a method of detecting edges in black and white images by smoothing the image using a Gaussian kernel and then observing the first and second difference equations and finding clear inflexion points, the points in the regions where are the change in greyscale intensity is greatest.

Another edge detection method is by using the Sobel Operator combined with a Gaussian kernel on the horizon image three pixels high. The Sobel Operator aims

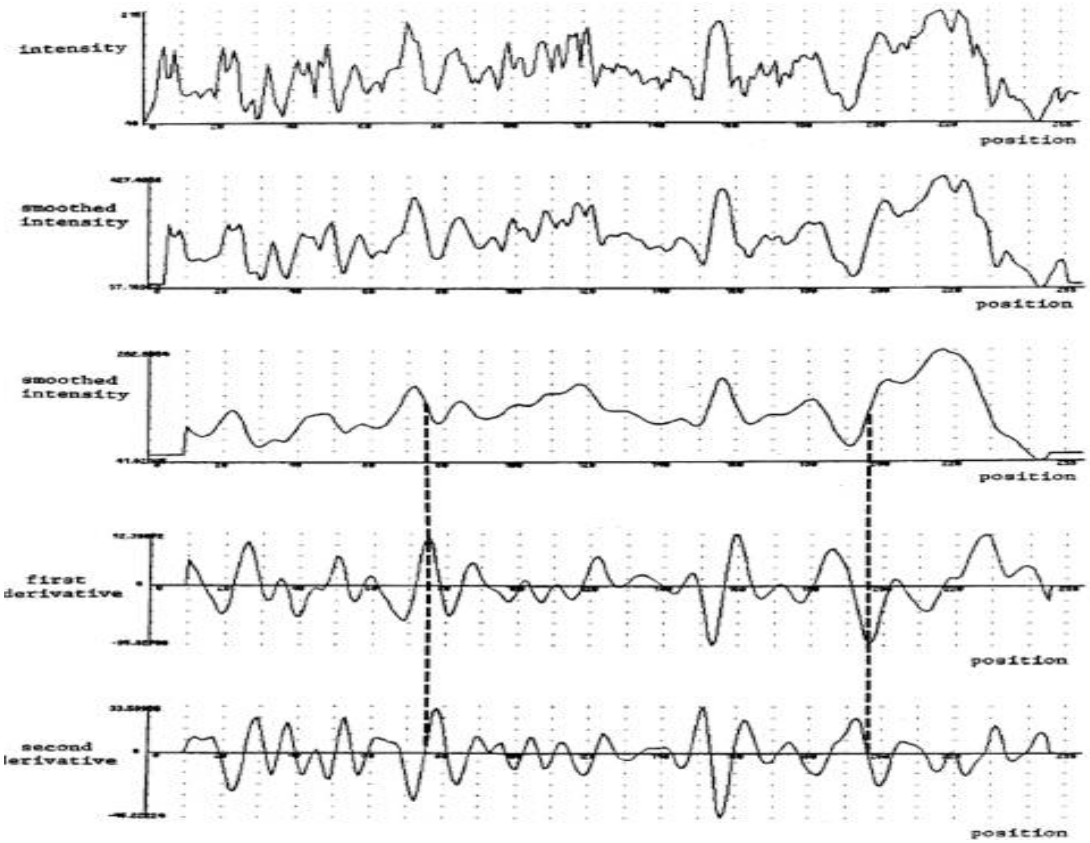


*Figure 1: Edge detection using differencing by Marr and Poggio.*

to give a measure of edge strength in the vertical and horizontal direction of an edge point. From there edge strength in arbitrary directions can be calculated. In particular the Scharr Operator given by:

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

for the horizontal direction and

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

for the vertical direction.

When edges are detected pixels can then be sorted by colour bands on the horizon. In the stereo vision technique when two robots receive different colour bands determining which colour bands belong to which image is a challenge. This problem translates partly into minimising the edit distance, in which there are dynamic programming algorithms that minimise the number of deletions needed to make the two colour bands equal. However one downside into these popular algorithms is that they include deletions and substitutions of colours to achieve this. Proposed below is the modified edit distance algorithm implemented in the stereo vision technique.

**Problem:** *Let $S$ be the set of all ordered colour bands. Given a sequence of two colour bands $A = a_1a_2a_3...a_k$ and $B = b_1b_2b_3...b_l, A, B \in S$, we wish to find $C = c_1c_2c_3...c_m \in S$, 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 $a_1a_2a_3...a_{i-1}$ and $b_1b_2b_3...b_{j-1}$ 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 $i \in (1,2...,k+1)$, $\delta(e[i,0]) = i, s(e[i,0]) = a_1a_2...a_{i-1}$ and $j \in (1,2...,l+1)$, $\delta(e[0,j]) = j, s(e[0,j]) = b_1b_2...b_{j-1}$. Define further $\beta(e[i,j])$ to be the largest consecutive set of recent deleted elements $b_jb_{j-1}...bj-c$ in determining $s(e[i,j])$ and $\alpha(e[i,j])$ to be the same $a_ja_{j-1}...aj-d$ in determining $s(e[i,j])$. $s(e[i,j])$ and $\delta(e[i,j])$ are determined as follows:

```
if (a_i ∈ α(e[i − 1, j − 1]) and b_j ∈ β(e[i − 1, j − 1])) {
    Dist_1 =δ(e[i − 1, j − 1]);
} else {
    Dist_1 =δ(e[i − 1, j − 1]) + 2;
}
if(a_i ∈ α(e[i − 1, j]) ) {
    Dist_2 =δ(e[i − 1, j]) - 1;
} else {
    Dist_2 =δ(e[i − 1, j]) + 1;
}
if(b_j ∈ β(e[i, j − 1])) {
    Dist_3 =δ(e[i, j − 1]) - 1;
} else {
    Dist_3 =δ(e[i, j − 1]) + 1;
}

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

if (δ(e[i, j]) == Dist_1) {
    if(Dist_1 ==δ(e[i − 1, j − 1])) {
        s(e[i, j]) = s(e[i − 1, j − 1]);
    } else {
        s(e[i, j]) = s(e[i − 1, j − 1]) without a_i and b_j;
    }
} else if(δ(e[i, j]) == Dist_2) {
    if(Dist_2 == δ(e[i − 1, j]) − 1) {
        s(e[i, j]) = s(e[i − 1, j]);
    } else {
        s(e[i, j]) = s(e[i − 1, j]) without a_i;
    }
} else {
    if(Dist_3 == δ(e[i, j − 1]) − 1) {
        s(e[i, j]) = s(e[i, j − 1]);
    } else {
        s(e[i, j]) = s(e[i, j − 1]) without b_j;
    }
}
```

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

# 3 Method

## 3.1 Stereo Vision Technique

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 $\Delta(Y) + \Delta(Y) + \Delta(Y) \geq 30$ is a good threshold condition.

4. Edge filtering is done to filter out noisy edges 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 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.

## 3.2 Profiling Technique

1. A ring of pixels representing the 360-degree horizon view from the center of the map is recorded in the robot.

2. The recorded ring of pixels is then binned into 28 regions, each region representing a certain corner of the map.

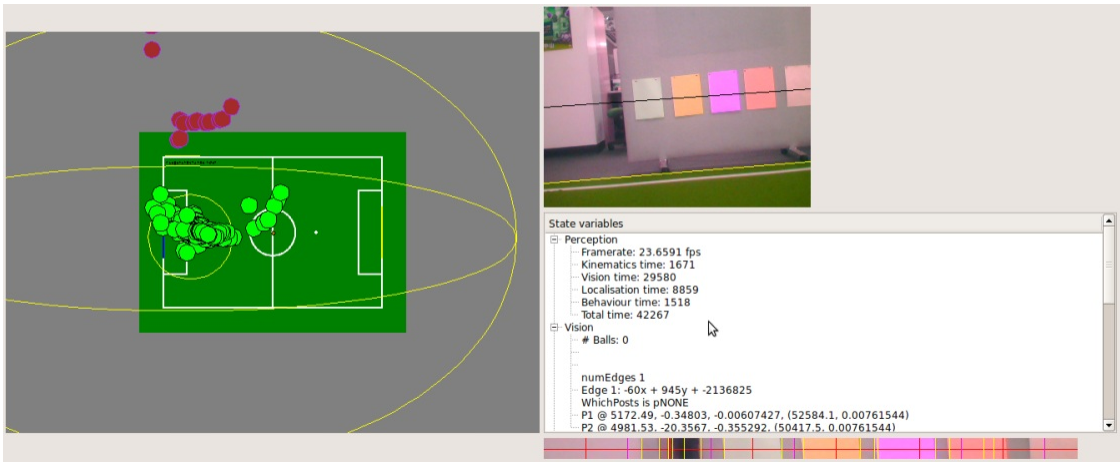


*Figure 2: Results from localising off coloured paper on a billboard.*

3. When the robot is on the map, its current horizon pixels is then compared to each of the 28 regions. A voting method is used so that the number of pixels in the current image belonging to the region is recorded is the number of votes.

4. The region with the most number of votes is the region that the robot is most likely facing.

# 4 Results

## 4.1 Stereo Vision 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 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 de-

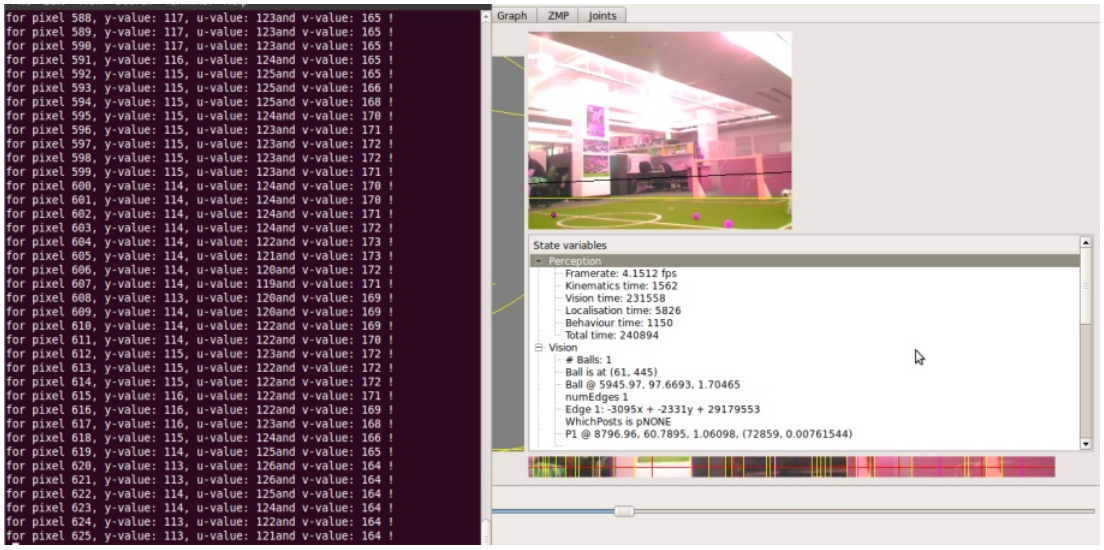


*Figure 3: Some YUV values of consecutive pixels shown with a sample image.*

tected 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.

### 4.2 Profiling Results

The regions method was accurate in detecting direction in objects with distinct colour differences, otherwise the instability of noise in the picture would be enough to make the direction point at different places over time. Profiling was most accurate for clear walls and was still accurate for wires and varying distances. But regions where the combinations of colour were similar often led to unstable readings, particularly if there were a wide range of colours in some regions.

## 5 Discussion

### 5.1 On the Stereo Vision Results

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 this was probably because

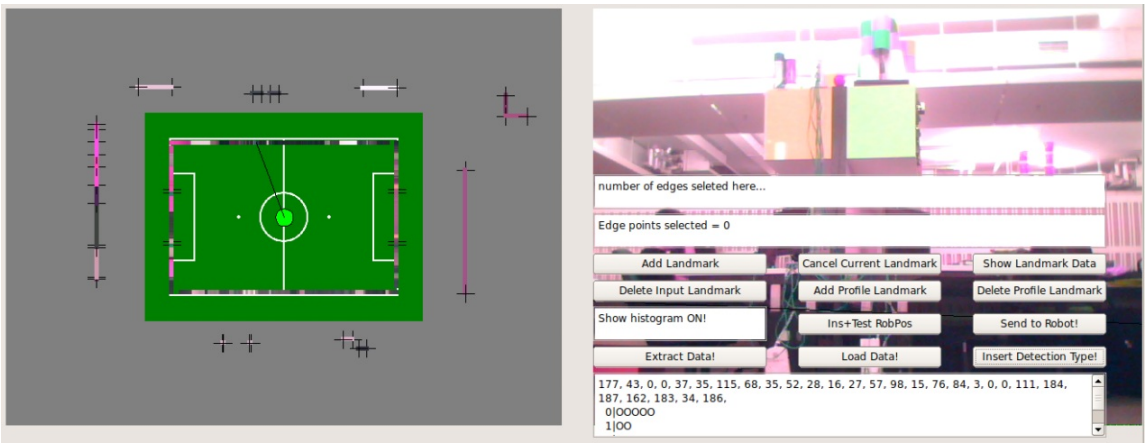


*Figure 4: Profiling results on the soccer field with sample image on the right.*

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.

### 5.2 On the Profiling Results

The claim was the profiling would be robust enough on the field to detect direction irrespective of current position, as long as it was on the field. This would naturally produce errors, because the actual horizon image distorts depending on the robot's true position on the map. Surprisingly, the error was rather insignificant in most cases. The more significant error came from grouping the ring of pixels into regions. The number and size of regions for optimal detection would vary, but 28 regions with 80 pixels in each seemed to be good regions. But regions with similar colours would confuse the robot, as the direction would randomly point from one similar region to another. This would especially be the case in symmetric maps, where regions would lose their distinctness and be rather ambiguous.

## 6 Conclusion

The Stereo Vision localisation, whilst potentially powerful, is not accurate enough for practical purposes. This is due to two main factors - the overall measurement error of most of the techniques implemented and the random and dynamic nature of environments producing too much noise in general. Perhaps with more thought and implementation a more accurate and reliable method could be found which can work, as time constraints were an issue in the ToR experience.

However the profiling results were more encouraging and seemed to be somewhat invariant towards field position, which was something of a surprise. More development would be needed to make it game-ready however. Some ideas include:

- Weighting different colours detected ('rewarding' the detection of contrasting colours).
- Looking at surrounding profiles, observing the histogram of region matches.
- Keep previous horizons profiles, so as to keep save states and use them in some way to improve computational times.

## 7 Reference


- Marr, D., and T. Poggio. A computational theory of human stereo vision. Proc. R. Soc. Lond. B Biol. Sci. 204: 301328, 1979.