Vision System for a Trax Robot

Donald G. Bailey
Institute of Information Sciences and Technology,
Massey University, Palmerston North, New Zealand
D.G.Bailey@massey.ac.nz

Abstract

An important component of a game playing robot is the vision system. The vision system is required to enable the robot to sense the current state of the game. This paper describes the tasks of the vision system for a Trax playing robot. Included in the discussion are the physical arrangement of camera and lighting, the consequent calibration of the system for lens and perspective distortion, determining when the playing area is free from obstruction (and the robot can make its move), and detecting the game pieces, or tiles for the game Trax.

Keywords: Trax, game playing, robotics, calibration, segmentation, template matching

1 Introduction

Computer games are nothing new—they have been around as long as people have had access to computers. What is novel with this project, however, is that we are building not just another computer game, but a robot that physically plays the game in much the same way that a human player would. Indeed, the goal of this project is to develop a system that seamlessly plays Trax against a human opponent.

Trax is a two-player abstract strategy game, invented in 1980 in New Zealand by David Smith [1,2]. It is played on a flat surface with square tiles with curved sections of black and white lines on one side, and straight lines on the other side (figure 1). One player plays black, and the other white, trying to form in their colour either a closed loop or line spanning from one side of the playing area to the other. Trax is a game of pure skill, since all of the pieces are identical and there is no luck involved.

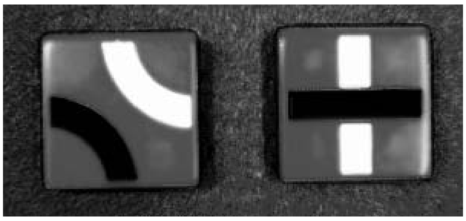

Figure 1: Trax tiles

In a turn, a player selects and plays a primary tile. This is played either side up against the tiles already in play, in such a way that the colours match where the paths join. As a result of playing the primary tile, there may be forced plays. Whenever two paths of the same colour enter a vacant space, a forced tile must be played that links the paths. Forced plays may also result in additional forced plays. Any such forced tile is played as part of the same turn as the primary tile. Therefore a turn may consist of several tiles being played. In fact it is these forced plays that give Trax its considerable strategic depth.

The structure of the game playing system is illustrated in figure 2. The computer player consists of three main components. The vision system, which is described in detail in this paper, enables the game playing robot to sense the game state. The game engine interprets the game state, and determines whether or not it is the robot's turn to play and what to move. The manipulator system enables the robot to make changes to the state of the game in the playing area [3].

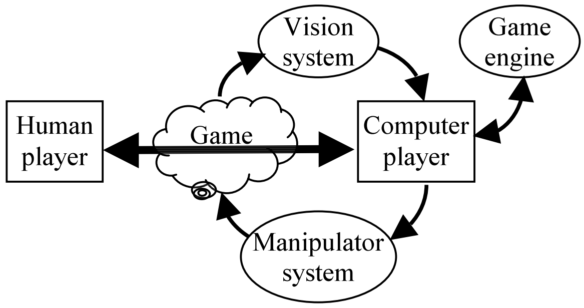

Figure 2: Components and interactions of the robotic game playing system, showing the interaction between the various modules.

The primary role of the vision system is to determine the current state of the game. For Trax this requires detecting the locations of the tiles within the field of view. These tiles are later sorted into two groups: those that are part of the playing area, and the pool of tiles used to make the moves.

However, since the tiles are moved by the manipulator system, it is important that the correspondence between the manipulator and vision system is determined. This is accomplished through a calibration procedure.

A third role is to determine when the human player has completed making their move. This is important because the human and robot share the same game area. The robot should therefore ensure that the human player is clear of the playing area before it moves the manipulator so as to reduce the chance of collisions.

The rest of this paper is structured as follows. Section 2 describes the physical arrangement of the image capture system, and the calibration between the vision system and the manipulator. Section 3 describes how the vision system is used to determine when the playing area is clear and the robot is free to move. Section 4 describes how the location and orientation of the Trax tiles is determined.

2 Physical Arrangement

The ideal location for the camera would be directly over the playing area with the sensor plane of the camera parallel with the playing area. The difficulty with this is finding a suitable location to place the lights. Both the tiles, and the tracks on the tiles have a strong specular reflection component. This means that lighting the playing area from near the camera will result in blind regions where there is strong reflections of the light source straight back into the camera.

This problem may be reduced to some extent by using diffuse illumination, although there is still a significant reduction in contrast when the light source is close to the camera. The reflectivity problem is made worse by the fact that the camera has a wide angle of view, requiring the lights to be a significant distance from the camera.

Lighting from a low angle avoids specular reflection, but introduces problems from the strong shadows that are cast around and between the tiles. These shadows are difficult to distinguish from the black tracks on the tiles.

The alternate arrangement is to view the playing area from one side, and to illuminate the game from the same side. This effectively moves the zone of strong specular reflection out of the field of view. Having the light source close to the camera significantly reduces the presence of shadows cast by the tiles from the field of view. This lighting arrangement is shown in figure 3.

The disadvantage of lighting from the side, however, is that it introduces significant perspective distortion. This results in a reduction in resolution at the edge of the playing area that is furtherest from the camera.

2.1 Calibration Procedure

Calibration is required to determine the correspondence between points seen by the input system, and points on the table. The simplest way of achieving this would be to have a grid positioned on the table, however this would require precise alignment when setting the system up. The alternative chosen was to move the manipulator to known positions within the field of view, and create a map that way.

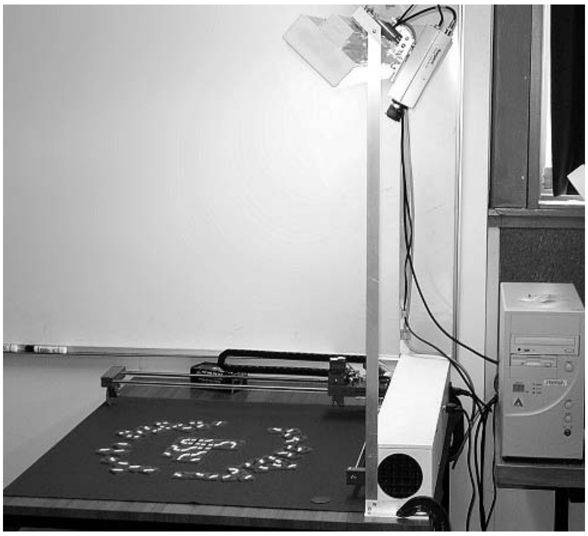

Figure 3: Arrangement of lights and camera

2.1.1 Target Detection

A simple target has been mounted on the back of the manipulator directly over the centre of rotation. The target was designed with 2 white quadrants and 2 black quadrants (see figure 4) so that it could be easily detected even if slightly distorted.

The target was detected using a 19x19 matched filter with weights as shown in figure 4. This filter has a maximum response when positioned over the centre of the target. To simplify the computation, the filter was decomposed into a 9x9 average, followed by a 4 element cross operator with 10 pixel spacing between the horizontal and vertical filters. The position of maximum response from this filter is assumed to be the target. The target location was refined to sub-pixel accuracy by taking the adjacent responses both horizontally and vertically, and fitting a parabola to the peak.

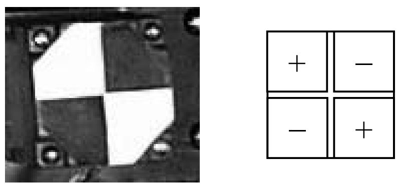

Figure 4: Target (left) and matched filter for detecting the target.

2.1.2 Correcting the Distortion

Viewing the playing area at a significant angle introduces significant perspective distortion. To have the camera within a reasonable distance of the playing area requires the use of a wide angle lens, and this introduces some barrel distortion. Both types of distortion are clearly seen in figure 5.

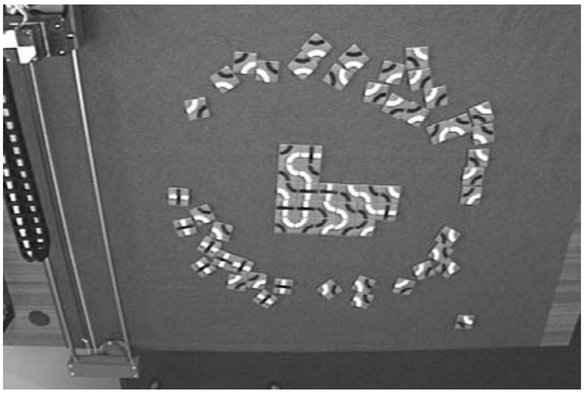

Figure 5: The camera's view of the playing area. Note significant perspective and barrel distortion.

To correct for the perspective distortion, a minimum of 4 points is required. The most convenient location of these points is in the corners of the area of interest. However the barrel distortion caused by the lens requires additional data from the image. This can be given by a further 4 calibration points in the centre of the sides.

The target is moved in a 3x3 grid pattern by the robot. This grid covers the full range of motion of the manipulator, and spans the full area of interest. The point in the centre of the playing area provides additional data that enables the centre of distortion to be found more accurately.

A modification of the camera calibration method described at last year's conference [4] is used to characterise and correct both the perspective and lens distortion. A simple perspective transformation preserves lines. Therefore any deviation from linearity of points that are meant to be in a line must be caused by lens distortion. First a parabola is fitted to each set of 3 co-linear points. The quadratic term characterises the degree of distortion, enabling both the centre of distortion and the radial distortion parameter to be determined. A simple first order radial lens distortion model is used

$$r_u = r_d(1 + \kappa r_d^2) \qquad (1)$$

where r_u and r_d are the distance from the centre of distortion in the undistorted and distorted images respectively. κ is the distortion parameter.

When corrected for this lens distortion, each parabola becomes a straight line. Representing the lines in homogenous coordinates, the perspective transformation is represented by a matrix multiplication, with 8 free parameters in the 3x3 homogenous transformation. These parameters are solved using least squares [4].

2.1.3 Moving Calibration to Table

The calibration procedure described above determines the correspondence between the target on the back of the robot and the pixels in the camera. However, the tiles are played on the table, which is about 20% further from the camera than the target. Therefore the calibration determined needs to be corrected for the height of the target above the table. Because the target and table planes are parallel, the transformation between them consists of a translation and a scale, as shown in figure 6.

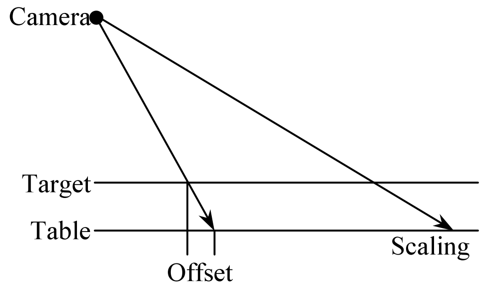

Figure 6: Scale and offset resulting from the target being above the table.

The position of the origin on the table is determined as follows: First the robot moves to the origin location, and a tile is placed under the gripper. This tile is located using the method described in section 4. Since the tile may have a slight offset (depending on how it was positioned), the gripper picks up the tile, and rotates it by 180 degrees about the origin. This way if the tile is offset from the origin, when it is rotated about the origin it will now be offset in the opposite direction. The mean of these two positions corresponds to the origin location.

The tile is then moved to the opposite corner of the playing area, and again located. Since the distance moved is known, this enables the scale factor between the target and the tiles to be determined.

2.2 Applying Correction

There are two ways in which this calibration information may be applied. Either the tiles can be located in the distorted image and correct the tile position and angle for the distortion, or the whole image can be transformed and all the tiles located in the undistorted image. The first approach requires far fewer computations while the second approach makes tile detection easier because all of the tiles are now the same size.

As processing time is not a significant issue in this application, the second approach was used. However, to speed the processing, a distortion index image is formed. For each pixel in the undistorted output image, the coordinates of the corresponding point in the distorted image are calculated. These coordinates are rounded to the nearest 1/16 pixel to enable bilinear interpolation to be used. Forming the distortion index image is only performed once and speeds the transformation of the captured image to correct for distortion. The corrected image is shown in figure 7.

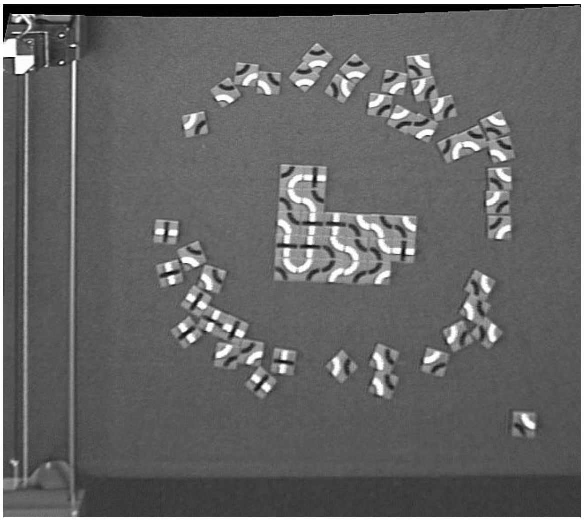

Figure 7: The image from figure 5 after correcting for perspective and lens distortion.

The resolution of the distortion corrected image is approximately 1.2 mm per pixel, although on the edge of the playing area away from the camera the true resolution is considerably less than this.

3 Determining Robot's Turn

A second task of the vision system is to determine when the human has finished making their move. This is accomplished by analysing the game state when the field of view is unobstructed to determine if the human player has completed a valid move.

The simplest approach to this is to detect any motion within the field of view while the computer is waiting for the human to complete their play. Since the motion does not need to be interpreted, a very simple motion detector suffices.

The absolute difference between successive frames captured about 1 second apart may be used to detect gross motion. For detecting motion, the undistorted images are able to be used, reducing the computation load. For each pixel, if the difference is greater that 40, that pixel is considered to have changed between the frames. A relatively large threshold was used because the black and white image was subject to interference from the colour subcarrier by the frame grabber. However the differences resulting from human motion within the field of view is almost always greater than this. If the number of pixels above the threshold is greater than 1200 (about 0.4% of the total pixels) then the playing area is considered obscured.

If the playing area is not obscured, the image is corrected for distortion and all of the tiles detected.

4 Tile Detection

Although the tiles have high contrast, the resolution is inadequate to reliably detect the edges of the tiles, especially when the tiles are placed adjacent to one another. The oblique camera angle and 6 mm tile thickness also means that the side of the tile toward the camera is also visible, particularly for the tiles that are furthest from the camera. Reliable detection of the location and orientation of the tiles is therefore based on detecting the patterns of black and white paths on the tiles.

When using a colour camera, the segmentation is straight forward, with good distinction between the green background, red tiles, and black and white tracks. With a black and white image, it is a little more complex with the tiles having a slightly lighter shade of grey than the background. This is exacerbated by the image being lit from the same side as the camera, as this creates an illumination gradiant preventing a simple global threshold to detect the paths on the tiles.

4.1 Image Segmentation

The first step is to distinguish between background and potential tiles. The tiles and background are distinguished on the basis of an activity index. The background is relatively uniform, whereas the tiles have significant activity with the black and white paths. This may be used to distinguish the background from other regions as follows:

- A 9x9 average filter is used to blur the image
- The absolute difference is taken between the blurred image and the original. If the difference is less than 5, the pixel is candidate background otherwise it is candidate tile.
- To reduce noise and other false classifications, another 9x9 filter is applied.
- The smoothed image thresholded at 50%, ie if there are fewer than 41 pixels initially segmented as background then that pixel is classified as background. Otherwise it is a tile pixel (or other object in the field of view, such as the robot).

Boundaries are not exact, but don't need to be. They are sufficient to reduce the search area when looking for tiles. The image is scanned again, this time thresholding for the black and white paths on the tiles. The white tracks are saturated, so a fixed threshold of 230 suffices to detect the white. The black tracks are more difficult. Since the illumination gradient varies vertically down the image, the median pixel value of those pixels classified as background in each row is determined. Anything that is less than 90% of this level, and was classified as tile is reclassified as black track. The resultant segmentation is shown in figure 8.

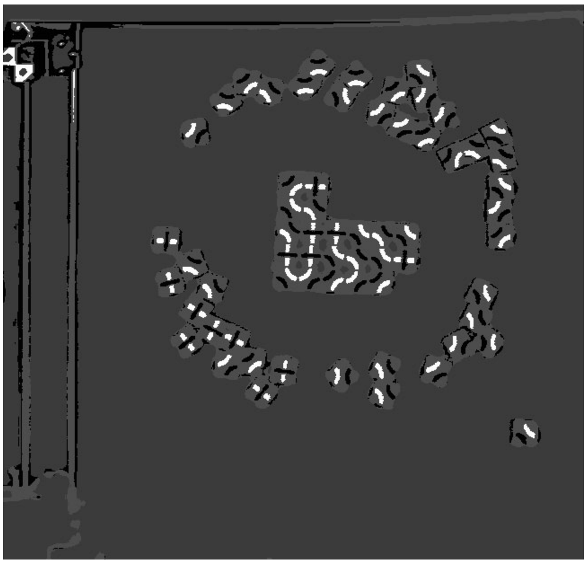

Figure 8: The results of segmenting the image of figure 7.

4.2 Tile Location

The image is then scanned to locate potential tiles using a template match. When looking from the centre of the tile, there are always 2 black pixels and 2 white pixels spaced 90 degrees apart at the ends of the tracks. Using this fact, 13 templates have been designed to cover all possible orientations of the 4 pixels at radius 9. These templates are shown in figure 9.

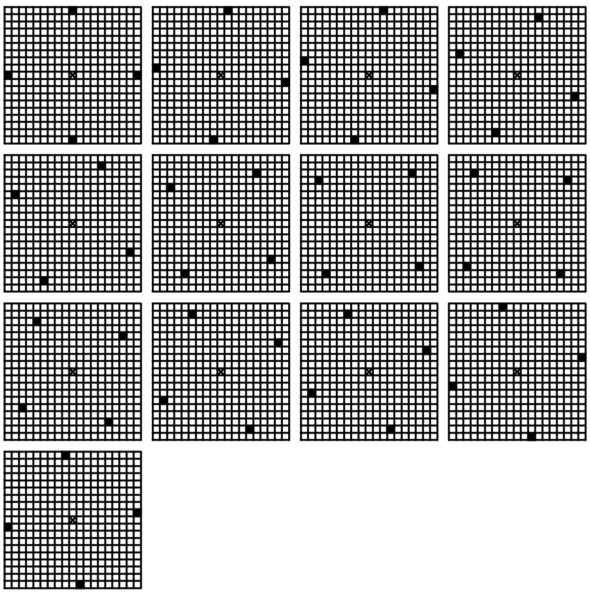

Figure 9: The set of 13 templates used for locating the centres of the tiles. The x is a possible tile centre if two of the ■ pixels are black and two white.

Since the centre of the tile cannot be on the background, this template is compared with every Rather than try to match this template to every pixel, only the pixels that were not labelled as background are considered. Also, since the tracks are several pixels thick, this will result in a cluster of candidate tile centres. Therefore to speed the search, only every second row and column in the image are scanned. If a match is found, the intermediate locations around that are then also checked. The particular template that was used is not recorded, as the tile orientation is determined more accurately in a later step.

The top left panel of figure 10 shows the candidate tile centres. Note that this approach is very effective at ignoring false black and white paths resulting from shadows or specular reflections.

For each cluster of candidate tile centres, the centre of gravity of the group is located. If the cluster has fewer than 4 or more than 25 connected candidate centres, the cluster is unlikely to be a tile and is discarded.

Occasionally, shadows between the tiles are detected as black track, resulting in some false tiles. These may be eliminated by checking tile signatures. From the centre of gravity, the labels at radius 9 pixels and 5 pixels are examined. The signatures for the detected tiles are shown in the top right of figure 10. There should be 8 changes (between B, W, and background) at radius 9, and either 4 (for a curve) or 8 (for a straight) changes at radius 5. The sequence of the changes is tested for validity. Any candidate centre that does not have a valid signature is discarded.

If the tile passes this test, the orientation and true centre of the tile is then estimated. For this all pixels with a radius of 12.5 pixels are considered, as shown in the bottom left panel of figure 10.

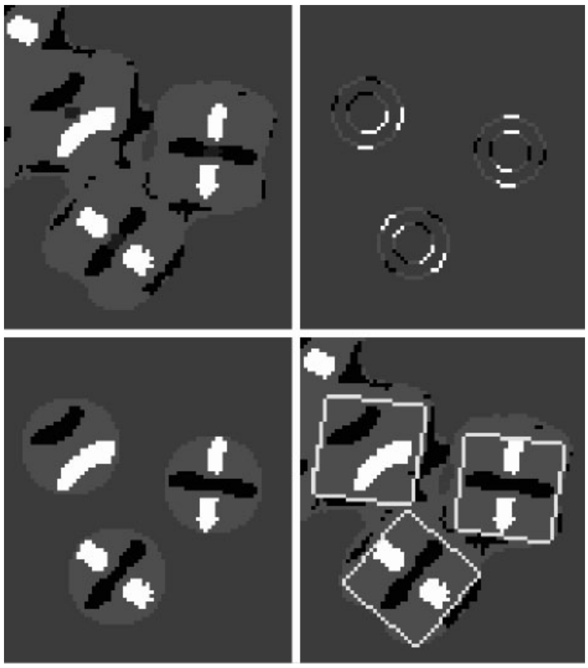

Figure 10: Processing to detect the tiles.
Top left: Template matching to find candidate tiles.
Top right: signatures of detected tiles.
Bottom left: selected regions for the detected tiles.
Bottom right: Orientation marked on original.

If it passes this test, all of the pixels within radius 12.5 pixels of candidate are considered. This radius ensures that if the detected track of adjacent touching tiles is appropriately segmented. The centre of gravity of all of the black, and all of the white pixels is measured. The true tile centre is midway between these two centres of gravity.

For curves, the tile orientation is given by the angle between the black and white and white centres of gravity (offset by 45 degrees). For straights, the orientation is found by examining the orientation of the black and white pixels (found using second moments). This is found by offsetting the black path orientation by 90 degrees and averaging it with that of the white path. The bottom right panel of figure 10 shows the detected tile location and orientation.

A final check on validity is performed. For curves, the path orientations should be approximately parallel, and for straights, approximately perpendicular. If this is not the case, the tile is discarded.

5 Summary

The vision tasks of a Trax playing robot have been described in detail. There are three important tasks of the vision system in this application.

The first is to calibrate the table positions with respect to the manipulator and camera. This is achieved by scanning the manipulator in a predetermined pattern and locating a target on manipulator. The calibration procedure can correct for both lens and perspective distortion. The calibration is moved from the manipulator plane to the table plane by translation and scaling by using the manipulator to control the location of a tile. Correcting for the distortions by warping the captured image takes less than a second.

A second task is to determine when the playing area is free from obstruction, allowing the robot to play. This is achieved by using a simple frame differencing scheme with one second between successive frames. In this way a simple motion detector is implemented.

The third task is to identify and locate the tiles within the field of view. For this and adaptive threshold is used to segment the image into background, tiles, black paths and white paths. The tile location is found by applying a simple template match to the paths to give candidate tile positions. These are then refined and verified by looking at the specific path configuration in the vicinity of the candidate centres. The use of such a multistep algorithm enables the tiles to be found robustly in an efficient manner.

6 References

[1] Smith, D., *How to Play Better Trax*, David Smith, Christchurch, (1983)

[2] Bailey, D.G., *Trax Strategy for Beginners*, 2nd edition, Donald Bailey, Palmerston North (1997)

[3] Bailey, D.G., Mercer, K.A., and Plaw, C. "Computer games: Out of the 'Box' and onto the Table", *Proceedings of the tenth Electronics New Zealand Conference, ENZCon'03*, pp 151-156 (2003)

[4] Bailey, D.G., "A new approach to lens distortion correction", *Proceedings Image and Vision Computing New Zealand 2002*, pp 59-64 (2002).