

Interpolation Methods for Global Vision Systems

Jacky Baltes and John Anderson

Department of Computer Science
University of Manitoba
Winnipeg, Canada
jacky@cs.umanitoba.ca
<http://www.cs.umanitoba.ca/~jacky>

Abstract. In 2004, the playing field size of the small sized league was significantly increased, posing new challenges for all teams. This paper describes extensions to our current video server software (Doraemon) to deal with these new challenges. It shows that a camera with a side view is a workable alternative to the more expensive approach of using multiple cameras. The paper discusses the camera calibration method used in Doraemon as well as an investigation into some common two-dimensional interpolation methods, as well a novel average gradient method. It also proves that (ignoring occluded parts of the playing field) it is possible to construct a realistic top down view of the playing field with a camera that only has a side view of the field.

1 Introduction

The small sized league (SSL) of the RoboCup competition is a very competitive league which has introduced many innovations and modifications into the rules in recent years. After a period of incremental changes to the field size (1997: 1.52m * 2.74m; 2001: 2.40m * 2.90m) the field size has been almost quadrupled for the 2004 competition. At the moment the rules have not been finalized but the proposed field size is 4.00m by 5.40m. Such a large change in the field size obviously led to much discussion of the pros and cons of this change. The authors of this paper hope that the larger field leads to more fundamental research in the SSL league and consider it necessary to move towards games of 11 vs 11 players in the future.

The greatest impact of the larger field size is on the global vision system. Whereas a single global vision camera with a wide angle lens mounted centrally on top of the playing field is sufficient for the 2002 field (2.40m by 2.90m), this is no longer true for the 2004 field.

Teams have suggested several solutions to this problem: (a) buy a better wide angle lens, (b) mount the camera at a higher position, and (c) mount several cameras over the playing field. The authors believe that a new model-based approach to global vision in the SSL is a viable alternative to these that will be beneficial to research. It is clear that few challenges remained for vision processing with a single overhead camera. However, there are still many open

problems when dealing with a side view of the field, not the least of which is to be able to compensate for occlusion of objects.

DORAEMON, the video server used by the Little Black Devils (LBD, named after the Winnipeg Rifles) from the University of Manitoba has always been an exception in the RoboCup competition. Instead of mounting the camera centrally overhead, the LBD have always used a side view of the playing field. The camera used by the LBD is an off-the-shelf camcorder *without a wide angle lens*, and DORAEMON is able to control our robots from any view of the playing field.

To be able to achieve this, Doraemon includes a sophisticated camera calibration routine based on the Tsai camera calibration. However, the larger playing field requires even better image processing and interpolation so that smaller geometric features can be detected and tracked successfully.

In this paper, the authors compare the accuracy and efficiency of several interpolation methods in the global vision robotic soccer domain. We also show a quasi-reconstructed view of the playing field obtained from a side angle.

2 Doraemon and the University of Manitoba Vision Kart

Figure 1 shows a picture of the University of Manitoba Vision Kart. The Vision Kart is a fully mobile global vision robotics setup. It contains a P4 2.4GHz small sized PC, a camera on a tripod, a side-mounted Infrared transmission system, and a wireless router. The Vision Kart can run off a car battery for about an hour and is an ideal platform for demonstrations and fund raising events for the LBD. Students usually use laptops with wireless cards to connect to the vision server and to control the robots.

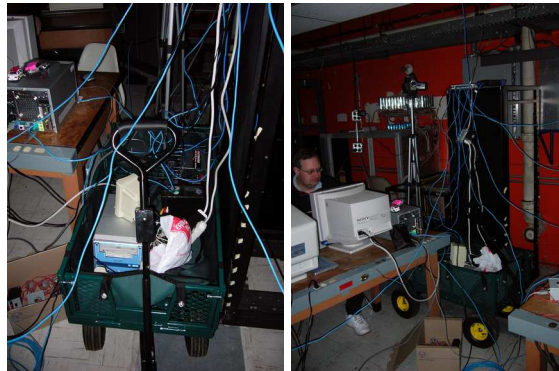


Fig. 1. University of Manitoba Vision Kart. A fully mobile SSL platform

Figure 2 shows the view of the camera in the vision kart overlooking the playing field. Clearly, the perspective distortion of this setup is much more extreme than in the case of an overhead camera. The top view shows the playing field with three robots. The bottom row shows a zoomed in view of the three robots.



Fig. 2. Overview from the camera on the vision kart (top) as well as a zoomed-in views of three robots.

From the views in Fig. 2, it is easy to see that colour features are hard to extract. For example, the small pink spot on the robot has been washed out to pure white because of the angle of the incoming light. We have therefore investigated alternative methods to use geometric features. For example, in [1] we describe a method that uses line segments of rectangular robots to determine the ID and orientation of robots. One problem of this approach is that most geometric features that may be used as feature points (e.g., area, aspect ratio, and angles) are greatly distorted by the low position of the camera.

3 Tsai Camera Calibration

DORAEMON’s camera calibration uses the well-established Tsai camera calibration [2] which is popular among computer vision researchers. It is also suitable for global vision in robotic soccer since it can compute the calibration from a single image.

The Tsai camera calibration computes six external parameters (x , y and z of the camera position as well as angles of roll, pitch and yaw) and six internal parameters (focal length, center of the camera lens, uncertainty factor S_x , and κ_1, κ_2 radial lens distortion parameters) of a camera using a set of calibration points.

Calibration points are points in the image with known world coordinates. In practice, Tsai calibration requires at least 15 calibration points. DORAEMON

uses a fast, robust and flexible method for extracting calibration points from the environment. A simple colored calibration carpet is used.

The user selects a colored rectangle and specifies the distance in the x and y direction between the centers of the rectangle. DORAEMON's calibration is iterative, so it can compensate for missing or misclassified calibration points. Even using a side view of the playing field, the calibration results in object errors of less than 1 cm.

4 Interpolation Algorithms

Interpolation has long been an active research area in computer vision, mathematics, and statistics. The problem can be stated as finding the value of a target function for points in between a supplied set of data points. Often the data points are assumed to be corrupted with noise, and smoothing must be performed first. This is often the case in computer vision applications such as global vision robotic soccer.

Figure 3 is a simple example of a one dimensional interpolation problem. Given the value of the target function at positions x_1, x_2, \dots, x_n , find the value of the target function for position x_t .

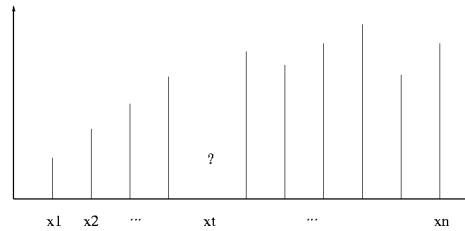


Fig. 3. One dimensional interpolation problem. Intuitively, one can see that the right side of the function is easier to interpolate than the left side.

By comparing the right and left half of the example interpolation problem, one can see that the right half is easier to interpolate. The data looks much more regular than the left side of the image.

4.1 Square Interpolation

The simplest interpolation algorithm is the square or pulse approximation, which results in a zero order interpolation of the target value as shown in Fig. 4.

This is identical to the one nearest neighbor algorithm (1-NN) algorithm where each sub-scanned pixel is assigned the value of its closest neighbor.

Although this algorithm is computationally very efficient, it does not lead to good results. For example, both the spot and the robot top have very jagged edges. This jaggedness also enhances false colour information in the image. For example, the yellow edges on the robot (which resulted from errors in the colour processing of the camera) are emphasized.

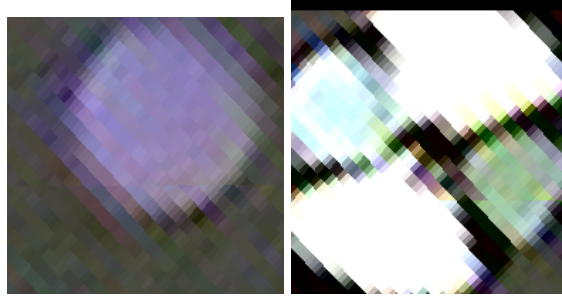


Fig. 4. Results of square pulse interpolation for the spots (left) and a robot (right)

4.2 Bilinear Interpolation

A slightly more complex interpolation method is the use of triangle instead of square approximation, also referred to as first order interpolation. In this case, the value of the target pixel is calculated along the line segment of its two neighbors.

The extension of this method to the two dimensional case (e.g., images) is slightly more complex. The problem is that four data points do *not* in general lie on an interpolation plane.

Two common methods to overcome this problem are: (a) piecewise linear interpolation, and (b) bilinear interpolation, as shown in Fig. 5.

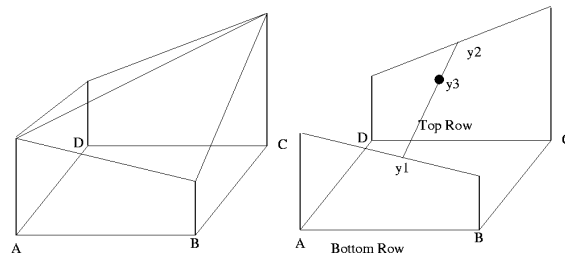


Fig. 5. Piecewise linear (left) and bi-linear interpolation of two dimensional data.

In the piecewise linear interpolation approach, the sample surface is split into two planes (ABC and ACD). In the top left of the image, points are interpolated using the plane ABC and in the bottom right using the plane ACD .

Using bilinear interpolation, points are first linearly interpolated along the rows/columns and the resulting points are then used to linearly interpolate the value of the target point. This bilinear interpolation is commonly used because it is computationally efficient and leads often to better results in practice than a square pulse interpolation. This can clearly be seen in the example pictures in Fig. 6. For example, the edges of spots are rounder and the shape of robot is also not as jagged as the square pulse interpolation.



Fig. 6. Results of bi-linear interpolation for the spots (left) and a robot (right)

4.3 Cubic B-Spline Interpolation

Another popular method of interpolation is to fit cubic b-splines to the data. In general, this method results in smoother interpolation of the data.

The main drawback is that it uses a larger neighborhood around the pixel to calculate the parameters of a cubic function. Because of computational constraints, the neighborhood is usually limited to a 4 by 4 pixel region.

The results of cubic b-spline interpolation are only slightly better than bi-linear interpolation, but larger interpolation neighborhoods are computationally much more expensive. Therefore, we restricted ourselves to 2 by 2 neighborhood interpolation functions.

4.4 Average gradient method

Because of the noise in the image, it is often necessary to apply blurring to pixels (e.g., to get a better estimate of the colour of region of pixels). The most popular method for blurring is to replace a pixel with the average or median of the pixels in its 4, 9, or 16 neighborhood.

However, blurring an image results in a loss of contrast around the edges of an object. Since these are important cues to determine the geometrical features that we use in determining the position, orientation, and identity of our robots, blurring is not appropriate in our case.

It is also computationally more efficient if the blurring and interpolation steps are combined.

We therefore designed and implemented a gradient blurring interpolation. The average gradient along the rows and columns are computed. The closest point to the target point is selected and the average gradient is applied at the starting point.

The resulting image of a spot and our robot are shown in Fig. 7. The result is similar to the bilinear interpolation and better than the square pulse interpolation.



Fig. 7. Results of our average gradient interpolation for the spots (left) and a robot (right)

5 Evaluation

One method of evaluating the scene recognition performance of a computer vision system is to reconstruct the scene from a different view point. Figure 8 shows the results of applying our interpolation method over the entire playing field. The resulting view is an overhead view isometric projection of the playing field using the original image as shown in Fig. 2.

The errors in the right edge of the playing field are due to errors in the camera calibration. In most cases, these errors are not significant, because the field is locally still consistent (i.e., the relative position of a robot and the ball is still correct), so robots can approach a ball to score, etc.

Although interpolating over the entire field is inefficient during a match, it nevertheless provides evidence of the accuracy that can be achieved with a solid camera calibration and reasonable interpolation methods.

Currently, no additional model knowledge has been used. The robots are assumed to be flat on the playing field. Although this is sufficient for detecting the position, orientation, and velocity of a robot, the quasi-overhead view reconstruction could be improved even further by correcting the image for the known height of robots that were found in the image.

We tested square pulse interpolation, bilinear interpolation, cubic b-splines interpolation as well as our own average gradient method on about 50 test images. Subjectively, we determined the accuracy of edges, area, and angles in the resulting image. In all cases, the performance of bilinear interpolation, cubic b-spline interpolation, and our own average gradient method was comparable and significantly better than that of the square pulse interpolation. The bi-linear interpolation is computationally the most efficient method, so we selected it as default interpolation method in DORAEMON.

One drawback of the interpolation method is that DORAEMON is unable to maintain 60 fields per second when controlling a field with 11 objects in it. The capture rate drops to 30 frames per second. We believe that better optimized interpolation routines or better object tracking can overcome this problem.

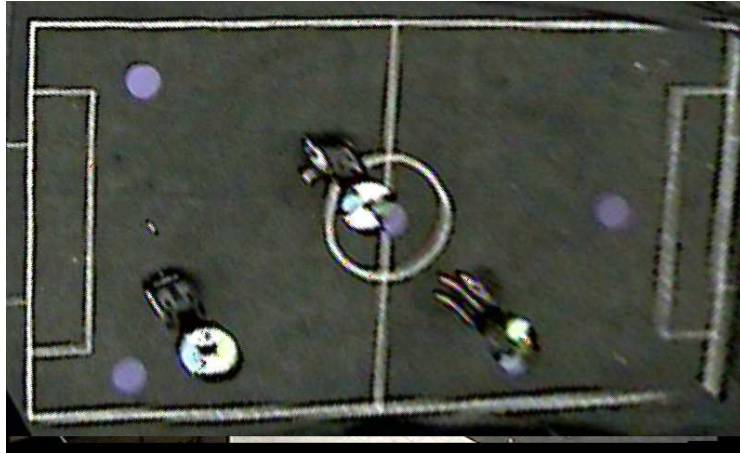


Fig. 8. This is *not* an image taken by an overhead camera, but rather a reconstructed overhead view. The original image is shown in Fig. 2. The image has been corrected for perspective distortion and our average gradient interpolation was used.

6 Conclusion

This paper argues for a more model-based approach to computer vision to tackle the problems posed by the new field size in the SSL league. The goal of our research is to use a single camera without a wide angle lens that is able to cover the new larger playing field. As a first step, we built on the existing camera calibration routines in DORAEMON to geometrically interpolate pixels. This interpolation is used to extract important geometric features such as the area, aspect ratio, and angles in a shape. These features are then used to determine the position, orientation, and identity of objects on the playing field.

This paper described a comparison of several interpolation methods including square pulse interpolation, bi-linear interpolation, cubic b-spline interpolation, and a novel average gradient method. In our evaluation, bi-linear interpolation had good results, but ran faster than cubic b-spline and average gradient interpolation. Square pulse interpolation has the worst results.

To show the effectiveness of the reconstruction, we showed the result of calculating the interpolation of the entire playing field. The resulting image will hopefully convince other researchers that overhead cameras are not the only solution to the vision problem in the SSL.

References

1. Jacky Baltes. Doraemon: Object orientation and id without additional markers. In *2nd IFAC Conference on Mechatronic Systems*. American Automatic Control Council, December 2002.
2. Roger Y. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 364–374, Miami Beach, FL, 1986.