A Panoramic Color Vision System for Following Ill-Structured Roads

Alan M. Zhang and Lindsay Kleeman
Intelligent Robotics Research Centre
Department of Electrical and Computer Systems Engineering
Monash University, Clayton, Victoria 3800, Australia
{alan.zhang, lindsay.kleeman}@eng.monash.edu.au

Abstract

The ability to follow ill-structured roads, such as footpaths, dirt tracks and corridors, is important for mobile robot navigation. This paper presents a panoramic colour vision based road following system for ill-structured roads. Roads are modeled with rapidly adapting 3D colour histograms and a simple yet generic geometric model. The computational complexity of the geometric model fitting stage has been significantly reduced compared to other works. A Kalman filter is used to smooth out any measurement noise. Results from experiments in tracking a footpath demonstrate the robustness of the system.

1 Introduction

Road following has been an active area of research for more than two decades. Helped by the prospect of commercialisation it continues to be an active area of research. Vision is the most common sensor used, although recent developments such as [Rasmussen, 2002] are trying to combine multiple sensing modalities for increased robustness. This paper focuses on road following using vision as the authors believe there is still much improvement to be made over existing algorithms.

Many road following systems are designed for well structured roads. These systems often rely on specific characteristics of the road such as colour, lane markings or edges [Kang and Jung, 2003], [Southall and Taylor, 2001]. Some of these systems, such as Carnegie Mellon University's Navlab system [Thorpe et al., 1991], has achieved remarkable results. Ill-structured road following on the other hand has received less attention, although the recent highly publicised DARPA Grand Challenge has renewed interests in this area. Ill-structured roads refer to those roads that lack readily and reliably extractable features used by the structured road following algorithms. They may not have lane markings, clearly distinguishable edges, consistent colour, or the ground may not be flat. These ill-structured roads include roads in the desert, as is the case in the DARAPA Grand Challenge, dirt/gravel tracks, footpaths, walkways, or even corridors. While some of these may be highly structured one should not expect to develop systems specific to each of them. A more general system that works in a variety of situations is more desirable.

The key to following ill-structured roads is adaptability. The description of a road must evolve as the road characteristics change. Crisman [Crisman and Thorpe, 1993] demonstrated the robustness achieved using this concept in their SCARF algorithm. SCARF resamples the colour statistics of the road on *every* frame. This allows for rapid changes in the appearance of the road (often caused by changing lighting conditions). Improvements to SCARF were made in [Ramström and Christensen, 2005]. However it suffers from high computational complexity and subsequently is restricted to operate on small images of size 61 by 50 pixels. This paper presents a road following system adapted from [Ramström and Christensen, 2005] to work with panoramic images and also makes significant improvements over their algorithm. In particular, the computational complexity has been reduced. A Kalman filter using a novel parameterisation of the road is also added to smooth out the noisy measurements. The paper is organised as follows: Section 2 describes the hardware setup; the major components of the road following system are presented in Section 3; Section 4 contains experimental results and discussion and Section 5 concludes the paper.

2 Hardware

The hardware setup is shown in Figure 1. The mobile platform used is a Pioneer 2AT robot. A colour webcam and panoramic mirror assembly is mounted at a height of approximately 1 meter above the ground. The assembly is vertical such that the horizon appears as a circle that is centred on the apex of the mirror. The mirror has the
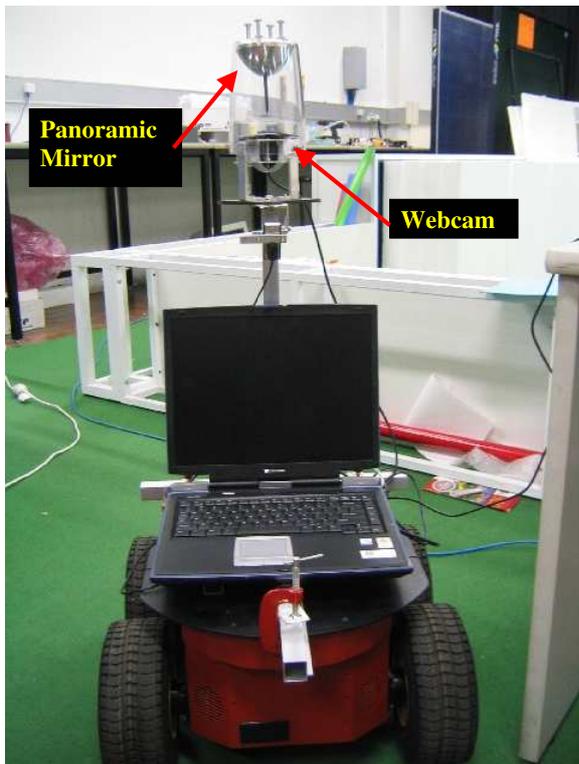
Figure 1: Hardware setup. Pioneer 2AT outdoor robot. Webcam and panoramic mirror assembly mounted on a pole approximately 1 meter from the ground. Laptop for data recording.

Figure 2: Overall system architecture.

profile presented in [Chahl and Srinivasan, 1999] (see Figure 3). This panoramic system has a large vertical field of view leading to a large portion of the image being occupied by the sky. This interferes with the automatic gain of the camera resulting in the road areas appearing too dark. A disk was therefore placed on top of the assembly to block out the sky. Image size is 320 by 240 pixels.

3   Road Following System Overview

Road tracking is achieved using the colour statistics of the scene. Earlier work in [Ramström and Christensen, 2005], [Crisman and Thorpe, 1988] has demonstrated the applicability of this approach at following ill-structured roads. Figure 2 shows the overall road tracking system architecture. It operates in the following manner. First the road location in the current frame is predicted from the previous frame. When the system is first initialised the robot is assumed to be on the road and pointing in the direction of the road. A "standard" road model having a width of a typical road is used as the initial road model. The on-road and off-road areas are sampled to build the colour models. The colour models are back-
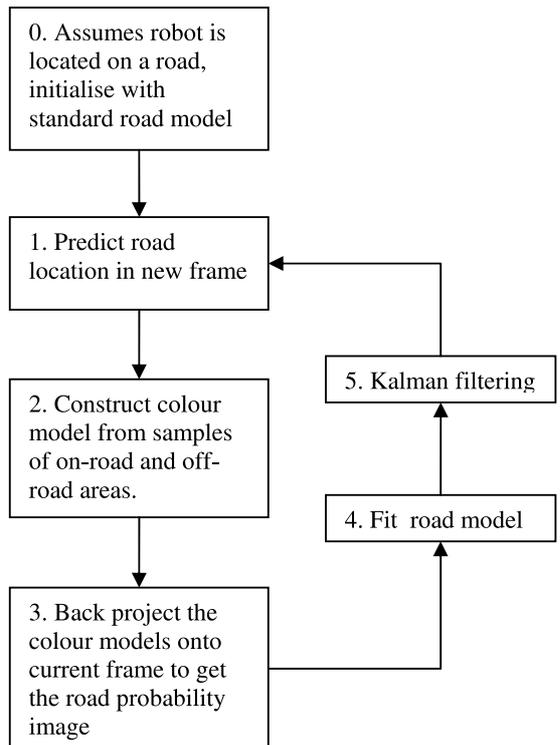
projected onto the current frame to give each pixel a probability of being a pixel on the road. A road model is then fitted to the road probability image using a global search. The result is fed into a Kalman filter to suppress noise. The final tracked road location is then used to predict the road location in the next frame. View of the road behind the robot was obstructed by a human operator who drove the robot for data gathering. So only the 180 degrees field of view forward of the robot is used for road following. Each component of the system is described in detail in the following sections.

3.1   Road Model

The road model was chosen to be as simple as possible to allow the system to operate on a wide variety of roads. The robot is not expected to operate at high speeds and this removes the need for accurate curvature estimation for the purpose of motion control. The most basic geometric property of a road is that it has left and right sides that are parallel. It is further assumed that the sides are also straight. This model is expected to impede the system's ability at handling curves. However, experiments showing successful navigation around some rather sharp bends demonstrate that this parallel straight lines model is adequate for low speed road following. To simplify matters, the ground is assumed
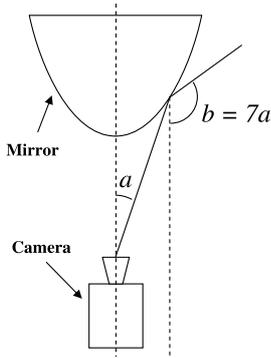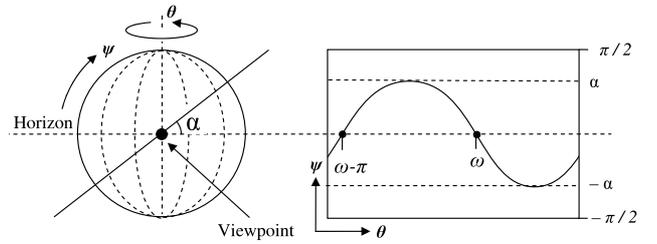
Figure 3: Panoramic mirror profile.

Figure 4: Projection of an infinitely long line on to a sphere. Diagram on the left shows the intersection between the sphere and the plane formed by the line and the viewpoint. The setup has been rotated about the vertical till the plane is viewed edge-on. The diagram on the right shows the "unwarped" image. The projected line is characterised by α the elevation angle of the plane and ω the vanishing point of the line.

to be locally flat and that the camera/mirror assembly remains vertical with respect to the ground plane. Consequently the horizon can be expected to remain at a fixed location in the image. Under these assumptions we analyse the projection of straight lines on the ground plane on to the panoramic image.

The panoramic mirror used here has the profile presented in [Chahl and Srinivasan, 1999] with a gain of 7. A nice property of this profile is that an image pixel's distance from the centre of the mirror is approximately proportional to the pixel's elevation angle, producing even pixel resolution for the whole range of elevation angles. Figure 3 illustrates this property. Unfortunately this property leads to the mirror not having a single effective viewpoint, i.e. the light rays reflecting off the mirror can not be modeled as having originated at a single point. However, for simplicity the setup can be approximated as *having* a single effective viewpoint. Successful experimental results in Section 4 verify the validity of this assumption.

Let's consider the projection of an infinitely long line on to a spherical coordinate system. The coordinates being θ the azimuth angle and ψ the elevation angle. The effective viewpoint and the line forms a plane. The intersection of this plane and the sphere is the projection of the line. Figure 4 illustrates this situation. On the left hand side the coordinate system is rotated about the vertical until the plane is viewed edge on. The "unwarped" coordinate is shown on the right. The projection can be expressed in spherical coordinates as follows:

$$\psi = tan^{-1}\left(tan(\alpha)sin(\theta - \omega)\right) \qquad (1)$$

where α is the elevation angle of the plane and ω is where the projected line intersects the horizon, i.e. the vanishing point. Therefore 2 parameters are needed to describe such a line: azimuth angle of the vanishing point ω and elevation of the plane α. A road model has a pair of parallel lines so only 3 parameters are needed. Let's refer to them as: α_l, α_r and ω_{lr}. Figure 5 illustrates these

parameters graphically in the coordinate system of the original camera image. The constant g in front of α_l and α_r is the constant of proportionality that converts elevation angle into pixel distance from the centre of the mirror. It is affected by the intrinsic parameters of the camera and the distance of the camera from the mirror [Chahl and Srinivasan, 1999].

Using the mirror's linear property between elevation angle and distance of the pixel from the centre of the mirror the panoramic image can be easily "unwarped" into a spherical coordinate representation. Figure 6 shows the original and unwarped images with projected road boundaries superimposed as a pair of red curves. Horizon is indicated by the magenta coloured dashed circle. Hence forth, I_{xy} refers to pixels in the original camera image and $I_{\theta\psi}$ refers to the same pixels expressed in spherical coordinates. The choice of spherical coordinates over cylindrical coordinates is motivated by the avoidance of singularities at ± 90 degrees elevation and that the mirror profile lends itself more naturally towards a spherical coordinate representation.

3.2 Road Probability Image

Before a road model can be fitted, each pixel needs to be assigned a probability of it being an on-road pixel given its colour: $P(R_{on}|I_{xy})$ where R_{on} is the class label indicating that the pixel is an on-road pixel. Applying Bayes' rule results in:

$$P(R_{on}|I_{xy}) = \frac{P(I_{xy}|R_{on})P(R_{on})}{P(I_{xy}|R_{on})P(R_{on}) + P(I_{xy}|R_{off})P(R_{off})} \qquad (2)$$

The priors are assumed to be uniform, i.e. $P(R_{on}) = P(R_{off}) = 0.5$. The quantities $P(I_{xy}|R_{on})$ and $P(I_{xy}|R_{off})$ can be calculated by modeling the colour statistics of on-road and off-road pixels. The samples
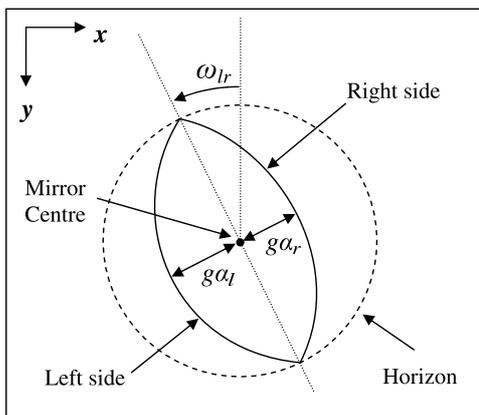
Figure 5: Road model and its parameters (α_l, α_r, ω_{lr}) in the coordinates of the original camera image. Constant g converts elevation angles α_l and α_r into pixel distance from the centre of the image.

used to train these colour models come from the predicted location of the road in the current frame. The predicted road model may mistakenly include off-road pixels as on-road pixels dues to errors in the predicted location or road curvature. To alleviate this problem safety margins are used as illustrated in Figure 6, where the yellow enclosure on the inside of the road is the on-road sample region and the two blue regions are the off-road sample regions. Experiments show that a safety margin of 3 pixels is sufficient. Note that pixels in the far distance are not used as on-road samples because of a possibly curved road.

Previous works suggest the use of Mixture of Gaussians (MoG) colour models trained with an EM algorithm [Crisman and Thorpe, 1988] [Ramström and Christensen, 2005]. While proven to be successful, it has a number of disadvantages. Firstly the training process is iterative requiring multiple passes through the pixels. Second, EM is a hill climb algorithm and thus may get entrapped in local maximas. As a consequence when on/off-road colour statistics change abruptly (possibly due to changes in lighting conditions) the mixture components may not be able to model these new points. There's also the choice of the number of Gaussians in mixture which needs to be determined experimentally. To avoid these complications 3D histograms are used to model colour statistics. Kristensen in [Kristensen, 20] concluded that the quality of classification between histogram and MoG are comparable if not slightly superior. It is simple to implement but has the drawback of requiring large amounts of storage. With the memory size of modern computers this is of little concern.

The histograms work in the raw RGB colour space. Other colour spaces, such as HSV, normalised RGB, YUV etc, and also 2D histograms were tested. But none
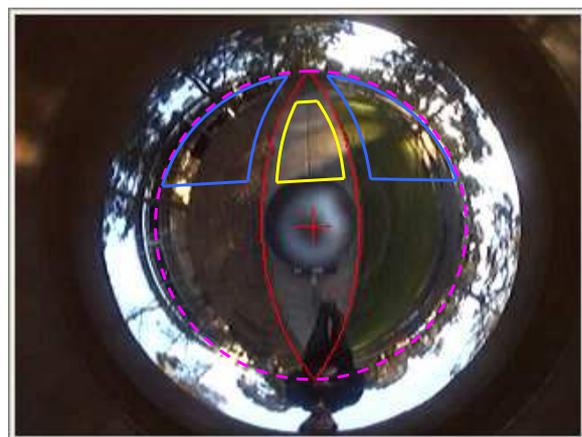
(a)

(b)

Figure 6: (a) Original image showing the predicted road location in red, on-road sample region in yellow, off-road sample region in blue and the horizon as a dashed magenta circle. (b) Unwarped image superimposed with the same road model as in (a)

performed better than raw RGB. Size of the histogram is 100 divisions for each of R, G and B dimensions. The probabilities $P(I_{xy}|R_{on})$ and $P(I_{xy}|R_{off})$ are modeled with two separate histograms. These are then smoothed with a symmetric Gaussian kernel to remove noise and fill gaps. The size of the kernel is 5 by 5 by 5 with a σ of 1.5. To speed up this operation a list of non-empty histogram entries is maintained during the histogram voting stage. Application of the smoothing kernel can then be restricted to only these non-empty cells. Equation 2 is now evaluated for every pixel in the current frame to obtain the road probability image. Equation 2 could be evaluated with a histogram division. But since the histogram is much larger than the image it is more efficient to evaluate Equation 2 for each pixel in the image. To prevent division by zero when no samples are present for a particular colour a small constant κ is introduced:

$$P(R_{on}|I_{xy}) = \frac{P(I_{xy}|R_{on})P(R_{on}) + \kappa}{P(I_{xy}|R_{on})P(R_{on}) + P(I_{xy}|R_{off})P(R_{off}) + 2\kappa}$$
(3)

Figure 7 shows the road probability image of Figure 6, where higher brightness represents higher probability of the pixel being an on-road pixel. Note that only the pixels below the horizon are processed in this stage. But to help the reader to get a better interpretation of the panoramic image the whole image is shown.

3.3 Model Fitting

It is suggested in [Ghurchian *et al.*, 2002] and [Ramström and Christensen, 2005] that the road probability image could be processed to obtain the locations of the road edges, followed by fitting a road model to these edge points. The alternative is to allow all pixels to contribute to the fitness of the model. The edge based method introduces errors through the road edge extraction process and has been shown to be less robust in [Ramström and Christensen, 2005]. For the method that uses all pixels, [Ramström and Christensen, 2005], [Kristensen, 2004] and [Crisman and Thorpe, 1993] suggest a weighted hough voting scheme. Each pixel will vote for all road models that contain that pixel as an on-road pixel with weight $P(R_{on}|I_{xy})$ and the rest of the models with weight $1 - P(R_{on}|I_{xy})$. The computational complexity is nS, where S is the size of the discretised road model parameter space and n is the number of image pixels below the horizon. It is because of this high computational cost that the images used in [Ramström and Christensen, 2005] and [Kristensen, 2004] are subsampled down to 61x50. Such small images place a limit on the achievable accuracy of the road following system. A novel model fitting method is presented here that reduces the computational complexity significantly. Since the pixels above the horizon are irrelevant to the fitting of the road model, from now on only those pixels below the horizon are considered.

First expressing the road model in spherical coordinates:

$$Left\ side: \ \psi_l(\theta, \omega_{lr}, \alpha_l) = tan^{-1}(tan(\alpha_l)sin(\theta - \omega_{lr}))$$
$$Right\ side: \ \psi_r(\theta, \omega_{lr}, \alpha_r) = tan^{-1}(tan(\alpha_r)sin(\theta - \omega_{lr}))$$
$$(4)$$

Building on the same weighted voting idea the fitness measure of a road model can be defined as:

$$f(\alpha_l, \alpha_r, \omega_{lr}) =$$
$$\sum_{\theta=0}^{\omega_{lr}} \left(\sum_{\psi=-\frac{\pi}{2}}^{\psi_l(\theta)} P(R_{on}|I_{\theta\psi}) + \sum_{\psi=\psi_l(\theta)}^{0} 1 - P(R_{on}|I_{\theta\psi}) \right)$$
$$+ \sum_{\theta=\omega_{lr}}^{\pi} \left(\sum_{\psi=-\frac{\pi}{2}}^{\psi_r(\theta)} P(R_{on}|I_{\theta\psi}) + \sum_{\psi=\psi_r(\theta)}^{0} 1 - P(R_{on}|I_{\theta\psi}) \right)$$
$$(5)$$

An intuitive interpretation of Equation 5, with reference to Figure 7, is that a good road model includes as much of the white pixels as possible within the road while at the same time excludes as much of the black area as possible.
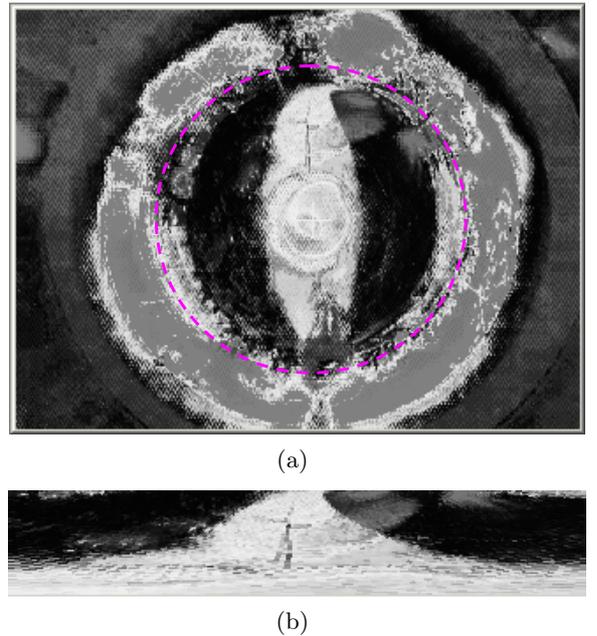

(a)

(b)

Figure 7: (a) Road probability of the image in Figure 6(a). (b) Expressed In spherical coordinates.

The fitness function is maximised using a global exhaustive search over a discretised parameter space. During the search, when ω_{lr} takes on a particular fixed value, the first half of Equation 5 is a function of α_l only and the second half is a function of α_r only. This means the two parts can be optimised independently. Therefore, the search space is only 2 dimensional.

The sums in Equation 5 can be evaluated efficiently using a column integral image of the road probability image in spherical coordinates. The integral image is the same size as the unwarped image $I_{\theta\psi}$ and its pixels are calculated as such:

$$J_{\theta\psi} = \sum_{\psi'=-\frac{\pi}{2}}^{\psi'=\psi} P(R_{on}|I_{\theta\psi'}) \qquad (6)$$

In a spherical coordinate system, the number of pixels at a particular elevation decreases as the elevation angle approaches vertical resulting in heavily interpolated pixels, as is evident near the bottom of the unwarped image in Figure 7(b). To counteract this effect each row in the unwarped image is given a weight. Lower weights are given to the highly interpolated rows. Using this integral image and the weighting of each row the fitness function can be evaluated by simply looking up the integral value in J for each point on the road boundary. The overall computational complexity is therefore $d_\theta S'$ where d_θ is the number of azimuth divisions of the unwarped image and S' is a 2 dimensional search space of the road

parameters. Using this efficient global search the system can operate at around 10fps on unwarped images of size 360 by 65 pixels on a 2.4GHz Pentium 4 machine. Road probability image computation and global search over all road models are by far the most computationally intensive components of the system, taking up on average 44ms and 61ms per frame respectively. Further improvements are possible. The search can be further reduced by concentrating the search around the predicted road location as opposed to an exhaustive global search. There are also improvements yet to be made for more efficient evaluation of the summations in Equation 5.

3.4 Kalman Filtering

The best road model is then passed as measurement into a Kalman filter to smooth the data according to vehicle dynamics and road curvature and width constraints. The state of the road is described by α_l, α_r and ω_{lr}. A popular description used in many existing algorithms [Ramström and Christensen, 2005], [Apostoloff and Zelinsky, 2003] is to track the position of the centre line, the width, and the orientation of the road. The advantage of the former state description is that it provides separate error estimates on the left and right side of the road. For example, when only the left side is measured reliably, the filter state should reflect this by having a smaller state error for the left side and a large one for the right. The robot controller might then decide to travel much closer to the left side instead of in the middle of the road. The constraint of a slowly changing road width is enforced by the filter process noise:

$$\alpha_l(t+1) = \alpha_l(t) + w_c + w_d$$
$$\alpha_r(t+1) = \alpha_r(t) + w_c - w_d \qquad (7)$$

where the normally distributed random variable w_c models the lateral shift of the robot with respect to the road, and w_d models the change in width of the road. The robot is a 4 wheel drive and uses skid steering. This compounded with varying road surface characteristic makes odometry unreliable. Odometric measurements are thus not included in the state prediction of the Kalman filter. Due to the slow speed of the robot ($\sim$0.5m/s) and a fast frame rate (10fps) the lack of accurate prediction step was not a problem.

4 Experimental Results

The test sequence was collected while a human operator drove the robot along a footpath. The video is 12 minutes at 10fps. For faster processing it has been down sampled by a factor of 2 to contain 3834 frames. Figure 8 shows some frames representative of the different conditions encountered. Figure 8(a) is the simplest case of a straight section of the footpath. Figure 8(b) shows
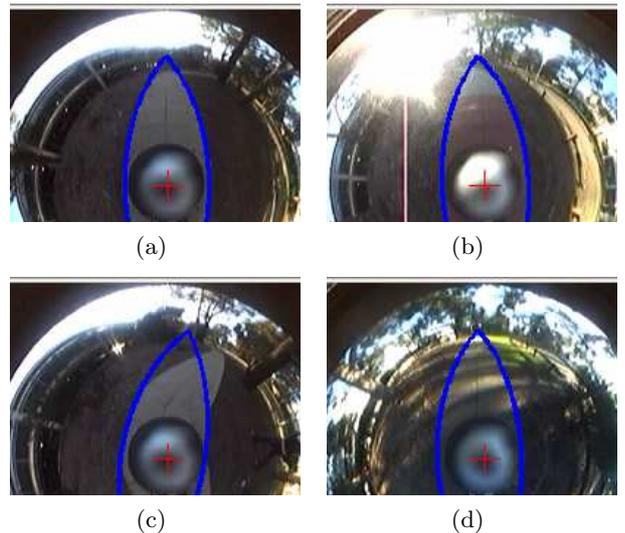

(a) (b)

(c) (d)

Figure 8: (a) A straight section of the road. (b) Abrupt road colour change from red to gray. (c) A sharp turn. (d) Streaking shadows from vegetation.

the road colour abruptly changing from red to gray. A rather sharp turn is present in Figure 8(c) and streaks of shadows are present in Figure 8(d). The tracked road is overlaid in blue.

It is difficult to create a measure that determines whether road tracking has failed. The fitness function in Equation 5 is not a good measure. For example, under some situations only one side of the road is measured reliably but this is good enough to keep track of the road and keeping the robot on the road. There are also situations where the road is not estimated correctly but not wrong enough to have lost track entirely and the system recovers after a while. So here we use a two tiered approach to detect tracking failure. First, if the variance of the road model parameters exceed certain thresholds tracking is considered a failure. Second, we will err on the side of caution and consider tracking failure if the estimated road location is judged to be wrong by visual inspection, even if the system recovers without interference. In the case of complete tracking failures the road model is manually reset to an "average road model" with a predefined width and orientation pointing straight ahead and in most cases the road is captured successfully thereafter.

For this data set there were only 2 tracking failures. First failure occurred from frame 754 to 873 due to sun glare obscuring the road. Frame 754 is shown in Figure 9(a). Second failure occurred between frames 2487 and 2586. This is caused by the footpath becoming a pedestrian crossing which does not have a significantly different colour distribution from its surround-

(a)   (b)

Figure 9: Two instances of tracking failures. (a) Frame 754, caused by sun glare obscuring the road ahead. (b) Frame 2487, caused by the pedestrian crossing not having a significantly different colour distribution from its surroundings.
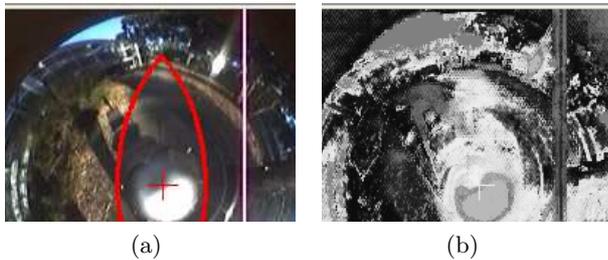

(a)   (b)

Figure 10: A difficult situation with a lot of shadows. (a) Red curves is the best fit road model, *not* the tracked road location. (b) Road probability image showing large amounts of noise.

ings. Frame 2487 is shown in Figure 9(b).

Figure 10(a) shows one of the more difficult situations. The image contains a large number of shadows, including the operator's and the robot's own shadows. The road probability image shown in Figure 10(b) is heavily corrupted by noise. Nevertheless, the road model fitting algorithm has correctly found the location of the road (drawn in red in Figure 10(a)), thus demonstrating its robustness. The video for this experiment is available at: http://www.ecse.monash.edu.au/centres/irrc/videos.php

5   Conclusion

This paper has presented a road following system based on the algorithm in [Ramström and Christensen, 2005] but adapted for a panoramic imaging system. Our algorithm differs from [Ramström and Christensen, 2005] in the use of a 3D histogram for colour modeling. The computational complexity of model fitting has been significantly reduced, allowing for the use of a large image. A Kalman filter has been added for smoother tracking. The filter states use a novel parameterisation of the road to allow for independent estimates on the tracking er-

rors of the left and right side of the road. Results have demonstrated a robust road tracking system capable of handling shadows, changing road colour and sharp road curvatures.

Acknowledgment

References

[Apostoloff and Zelinsky, 2003] N. E. Apostoloff and A. Zelinsky. Robust vision based lane tracking using multiple cues and particle filtering. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, Columbus, OH, USA, 2003.

[Chahl and Srinivasan, 1999] J. S. Chahl and M. V. Srinivasan. Panoramic vision system for imaging, ranging and navigation in three dimensions. In *Proceedings of the International Conference on Field and Service Robotics FSA'99*, Pennsylvania, USA, 1999.

[Crisman and Thorpe, 1988] Jill D. Crisman and Charles E. Thorpe. Color vision for road following. In *SPIE Conference on Mobile Robots*, Cambridge, Massachusetts, November 1988.

[Crisman and Thorpe, 1993] J. Crisman and C. Thorpe. Scarf: A colour vision system that tracks roads. February 1993.

[Ghurchian *et al.*, 2002] R. Ghurchian, T. Takahashi, Z. D. Wang, and E. Nakano. On robot self-navigation in outdoor environments by color image processing. In *Seventh International Conference on Control, Automation, Robotics And Vision (ICARCV'02)*, Singapore, December 2002.

[Kang and Jung, 2003] Dong Joong Kang and Mun-Ho Jung. Road lane segmentation using dynamic programming for active safety vehicles. *Pattern Recognition Letters*, 24(16):3177–3185, 2003.

[Kristensen, 2004] Daniel Kristensen. Autonomous road following. Master's thesis, Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Stockholm, Sweden, 2004.

[Ramström and Christensen, 2005] O. Ramström and H. I. Christensen. A method for following of unmarked roads. In *Intelligent Vehicles '05*, pages 650–655, Las Vegas, NV, June 2005. IEEE.

[Rasmussen, 2002] Christopher Rasmussen. Combining laser range, color, and texture cues for autonomous road following. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, Washington, DC, May 2002. IEEE.

[Southall and Taylor, 2001] Ben Southall and C. J. Taylor. Stochastic road shape estimation. In *Proceedings of the 2001 IEEE International Conference on Computer Vision*, pages 205–212, Vancouver, Canada, 2001.

[Thorpe *et al.*, 1991] C. E. Thorpe, M. H. Hebert, T. Kanade, and S. A. Shafer. Toward autonomous driving: The cmu navlab. *IEEE Expert*, pages 44–52, August 1991.