

Cylinder Sweep: Fisheye Images into a Bird's-eye View

Gabriel Hartmann and Reinhard Klette

The *.enpeda..* Project, The University of Auckland
Tamaki Innovation Campus, Auckland, New Zealand
gabriel.hartmann@gmail.com

Abstract. One focus of driver assistance systems is on the improvement of the driver's visual perception of the environment. This is often done by augmenting the limited field of view of a driver through the use of strategically placed cameras. Because of limitations on the amount of time a driver may look away from the road, it is desirable to limit the number of views a driver must analyse. Aesthetic and monetary considerations likewise place a priority on a limited number of cameras. This paper proposes an image stitching technique based on an extension of plane sweeping which provides drivers with a single omnidirectional bird's-eye view of the environment. This image is derived from input from cameras with fisheye lenses whose wide field of view limits the number of cameras required to guarantee an omnidirectional result. The system's resultant images improve upon previous birds-eye view simulation efforts in both elimination of seams between cameras, and in effective area of view.

Keywords: Image stitching, fisheye camera, bird's-eye view, vision-based driver assistance

1 Introduction

The value of providing vehicle drivers with more views of their environment than just two eyes can provide has been acknowledged since at least 1904 when rear-view mirrors were referenced as being used in horse-drawn carriages. Today the availability of inexpensive digital video cameras allows more views, from different angles than mirrors alone can provide to a driver, but the amount of time a driver has to consult these images has drastically shrunk as vehicle speeds have increased. An ideal solution would be a single omnidirectional image which can be easily interpreted by any driver in a short period of time. A bird's-eye view of a vehicle and its environment fulfils both these criteria. Direct production of this perspective via a flying vehicle or a camera on a boom is impractical for a number of apparent reasons. Therefore it is desirable to create a view which apparently comes from a position several meters above a vehicle, but which is produced by the combination of images from cameras which are actually mounted on the vehicle.

Several attempts have been made to construct such a system. Perhaps the most successful effort, produced by [Liu et al. 2008] produced nearly seamless

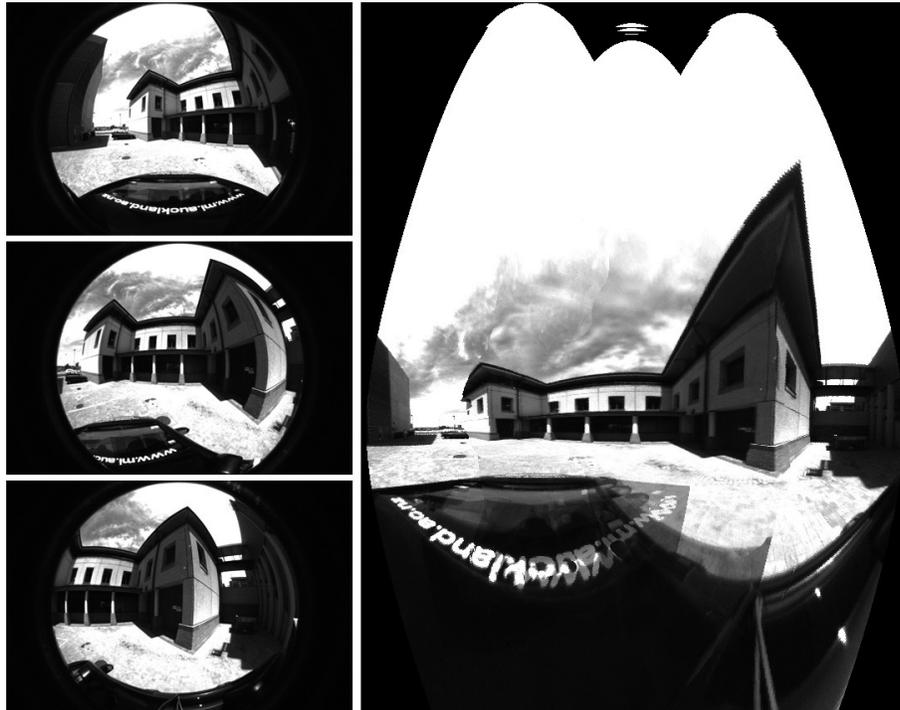


Fig. 1. *Left:* Three fisheye input images, parallel to the axis of vehicle motion; 45° clockwise and 90° clockwise relative to the axis of vehicle motion respectively; shown top to bottom. *Right:* stitched cylindrical image.

bird's-eye view images that were limited to displaying objects on or very near the ground plane, due to a reliance on homography with the ground plane. [Gehrig et al. 2008] were able to produce a 3D perception driver assistance system using fisheye lenses, but these cameras were still oriented in the traditional, roughly parallel configuration, typical of stereo systems. This system was focused on forward motion and analysis of the environment ahead of the vehicle, and so aimed for a field of view no greater than 150° .

This paper describes a novel technique for stitching wide field-of-view images together which is capable of producing 360° panoramic cylindrical images (see Fig. 1). This method extends the plane-sweep technique employed by [Kang et al. 2004] to produce seamlessly stitched images, by sweeping right circular cylinders instead of planes, and working on video sequences instead of static images.

[Liu et al. 2008] have previously developed a cylinder sweeping system for reproducing stereo reconstructions of environments. Their focus on producing 3D depth information relied on careful physical rotation of a camera or cameras around a central point to produce cylindrical images. The technique introduced

here does not produce cylinder models through physical motion. Instead, the cylinders are virtual objects upon which real images are projected. Two advantages of this approach are the possibility of static (relative to the test vehicle) placement of cameras as well as arbitrary arrangement of cameras around the periphery of the test vehicle. However, with this technique 3D information is only produced as a byproduct of the 2D cylindrical stitching process and is of only limited quality when compared to the results in [Liu et al. 2008] and [Kang et al. 2004].

The stitched sequence frames simulate the output of an omnidirectional video camera which produces cylindrical images. In order to produce the final bird's-eye view illusion desired, these cylindrical images are subsequently distorted as illustrated below in Fig. 5.

The paper is structured as follows. Section 2 discusses the applied camera calibration procedure. Section 3 informs about the proposed image stitching method, including the cylinder sweeping method and a brief report on experiments. Section 4 concludes.

2 Camera Calibration

In order to stitch any two images together it is necessary to determine which portions of the two images are common to both. Often this is accomplished by a process of feature matching, or through careful determination of the position and distortion characteristics of the cameras and lenses in question, or through some combination of the two techniques. The approach adopted here does not use any feature matching to align images, but instead relies solely on characterizing what are commonly referred to as the intrinsic and extrinsic properties of the cameras using the Camera Calibration Toolbox for Matlab [Bouget 2010]. The intrinsic characteristics refer to the distortions that cameras and lenses introduce to an image, and the extrinsic properties refer to the position and orientation of the cameras relative to each other.

Fisheye lenses offer the advantage of extremely wide fields of view, in this case (i.e. that of the used lenses) approximately 185° , which promise the ability to produce panoramic images with a minimum number of cameras. See Fig. 2 for an example of an input image. However the high degree of radial distortion introduced by fisheye lenses introduces nontrivial difficulties to the image stitching process. In order to make practical use of the information provided by the fisheye cameras, it is necessary to remove the distortion introduced by those cameras. The Camera Calibration Toolbox is able to effectively model and thus allow for the removal of this distortion. However the system's design, which attempts to simulate an ideal pinhole camera, necessarily limits its applicability to imaging systems with less than 180° fields of view, thus eliminating some of the theoretical benefit of using fisheye lenses.

The Camera Calibration Toolbox models fisheye images with four radial distortion coefficients κ_1 , κ_2 , κ_3 , and κ_4 . Let θ be the angle between the optical axis and a ray whose origin is at the focal point and extends towards a point in

an undistorted pinhole image. The radially distorted angle is

$$\theta_d = \theta(1 + \kappa_1\theta^2 + \kappa_2\theta^4 + \kappa_3\theta^6 + \kappa_4\theta^8) \quad (1)$$

Radius r is the length of the vector \mathbf{p} emanating from the optical centre and extending to the the point of interest in the pinhole image's optical plane. The vector

$$\mathbf{p}_d = \left(\frac{\theta_d}{r}\right)\mathbf{p} \quad (2)$$

is the vector \mathbf{p} radially distorted. Equations (1) and (2) provide a lookup based mechanism for producing undistorted images, based on the radial distortion coefficients determined by the toolbox. Given a position in an undistorted pinhole image, these formulas provide the position after distortion. In this case the result is the relevant position in the original fisheye image.

In the test system evaluated in this paper, cameras are never positioned with optical axes greater than 90° apart, simulating a scenario in which four cameras, stationed at 90° intervals around the periphery of a car can produce a complete panorama. This configuration does not require use of the cameras' full field of view, and the use of 162° (90% of π) fields of view was found to be more than sufficient. The right image in Fig. 2 shows the result of an undistortion of the left image with a 162° field of view constraint in both the vertical and horizontal axes.

The extrinsic properties of the system were also determined using The Camera Calibration Toolbox. The relative positions of cameras are provided in the form of Rodrigues rotation and translation arguments [Rodrigues 1840]. This information allows 3D positions in one camera's coordinate system to be converted to coordinates in any other camera's coordinate system.



Fig. 2. *Left:* original fisheye image. *Right:* planar perspective undistorted image.

3 Image Stitching

Our bird’s-eye views are generated by applying a cylinder sweep which builds upon the ideas of a previously proposed plane sweep.

3.1 Plane Sweep

Given the intrinsic and extrinsic properties of two or more cameras with overlapping fields of view and roughly parallel optical axes it should, in principle, be possible to produce a single seamless image from the two input images. Plane sweep was originally used as an early method for producing dense depth maps, but was adapted by [Kang et al. 2004] for stitching images. From the plane sweep perspective, the stitching process is one of systematically resolving disagreements between the cameras as to the colours of pixels in the overlapping region. This is done through the production of *disparity space images* (DSI).

Each DSI is a planar image parallel to the camera image planes at a given distance. For each pixel of the DSI, the cameras are consulted as to what they believe the colour of the pixel should be according to a perspective projection. A grayscale value is assigned based on the level of their agreement. If there is perfect agreement between the two cameras, the DSI pixel is coloured white; for complete disagreement, black. This process is repeated at a number of speculative distances from the camera. When a complete set of DSIs has been created, pixels in the overlapping region are painted with that colour about which the two cameras were determined to have the highest degree of agreement. This process works well for cameras with low degrees of radial distortion, and for situations which have no expectation of omnidirectional results.

3.2 Cylinder Sweep

As sweeping planes was originally conceived as a process for determining depth information, so sweeping cylinders was also originally employed to produce dense depth information for panoramic scenes by [Liu et al. 2008]. Here we apply the use of cylinder sweep to the problem of stitching together cylindrical panoramas from fisheye images. See Fig. 1, left, for an example of three input images.

Cylinders, as sweeping surfaces, offer two advantages over planes when attempting to produce full field of view panoramic images from images produced via fisheye lenses. First, whereas rays projected from a point lying outside a plane, through a plane, can only approach asymptotically a 180° field of view in both dimensions, rays projected from a point inside a cylinder can produce a 360° field of view in one dimension, and match the performance of a plane in the other dimension. Second, the curved geometry of a cylinder in one dimension much more closely approximates the geometry of a fisheye lens, thus avoiding severe undistortion effects (see Fig. 2) in at least in one dimension.

All the circular right cylinders created in the process of producing a DSI set share a common axis. This axis, corresponding to the optical axis of the simulated bird’s eye view camera, is perpendicular to the ground plane and intersects the

centre of the roof of the test vehicle. The radii of the cylinders are defined by the following formula:

$$R_c = R_{min} + i^2 \cdot S \quad (3)$$

R_c is the radius for a cylinder with index i , minimum radius R_{min} and radius step size S .

The minimum cylinder radius employed in this test system is 2 m, a distance which comfortably encompasses the entire test vehicle. Radii for each subsequent cylinder in a DSI set is increased quadratically in accordance with the inherent depth accuracy constraints of a fixed baseline system. The maximum radius for a cylinder is ≈ 80 m, a distance at which the limits of the camera baselines disallows further comparison of camera frames.

Each cylindrical DSI in a set is populated with colours in much the same way as in a plane sweep scenario. After correcting for radial distortion, optical center position, and focal length variations, as described in Section 2, the frame captured by each camera is projected onto a single shared cylinder. The DSI is created by measuring the levels of agreement between all relevant cameras for each pixel position on that cylinder. Each pixel in the overlapping region is assigned a greyscale value encoding the level of agreement between cameras. This process is repeated approximately 30 times for each set of synchronously captured frames with cylinders of quadratically increasing radius.

So for every set of frames, a set of cylindrical DSIs is created. The DSI set is used to determine, for each pixel in an overlapping region, the best colour for that pixel. Recall that cylindrical DSIs are produced by projecting idealized pinhole images onto cylinders with a shared axis. This means that pixel coordinates in the DSIs correspond to locations along a ray beginning at the intersection of the cylindrical axis and the roof of the car. That is, all pixels in a DSI set sharing the same coordinates are guaranteed to lie along a ray emanating from the focal point of the simulated omnidirectional camera, sampled at quadratically increasing distances. In order to find the best colour for a pixel position (x, y) in an overlapping region it is only necessary to find which position along a ray enjoys the highest level of agreement between projections, or more simply, which DSI pixel is whitest at (x, y) .

Once a position along a ray has been determined to have the highest level of agreement the original colour values from the various projected frames contributing to this position are again consulted to determine the final colour for a position in an overlapping region. Three different approaches were considered for determining the final colour. In the first case the average of the contributing colours is taken. In the second, the colour from the camera which saw the position straightest on is taken. More precisely, that camera which had the smallest angle between its optical axis and the ray extending from its focal point to the position in question was determined to have the authoritative colour.

The third approach is a hybrid of the first two and is defined by

$$W = \frac{|C_{avg} - C_s|}{C_{max}} \quad \text{and} \quad C_{stitch} = W(C_s) + (1 - W)(C_{avg}) \quad (4)$$

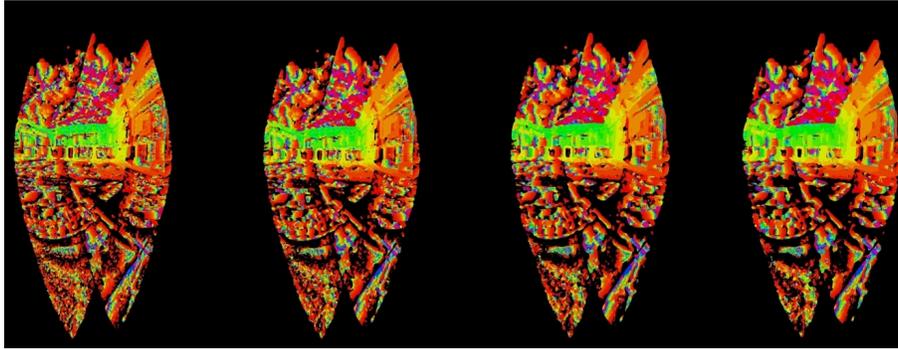


Fig. 3. Images resulting from 5x5, 7x7, 9x9 and 11x11 cost function windows respectively (Colours represent distances: red is closest, violet is farthest away)

W is the weighting factor associated with the colour value contributed by the camera with straightest view of a given pixel position. C_{avg} is the average colour of all contributing colours for a position and C_{max} is the maximum possible colour value. C_s is the colour contributed by the camera with the smallest difference in angle between the optical axis and the ray from its focal point to the pixel position in question. C_{stitch} is the colour of a pixel in the final stitched image.

A weight is introduced which is the result of a comparison between the colour contributed by the camera with the straightest view (hereafter referred to as the straight camera) and the average colour of all cameras viewing a position. If the average colour from all cameras for a given pixel is in accord with that colour contributed by the straight camera, then the straight camera's colour is given no special weight. However as the average colour diverges from the straight camera's, more weight is given to the straight camera's colour.

Areas in the final cylindrical image that are visible to only one camera are problematic. The cylinder sweep technique relies upon comparing the perspectives of at least two cameras over a series of increasing radii. When there is only one camera viewing a particular position this approach breaks down. An ad-hoc approach, which is in need of improvement, is used here. In each DSI set, that cylindrical image which is whitest, that is, that image which has the highest level of agreement among all pixels is determined to be a default layer for that frame. If a particular pixel is only viewed by a single camera, the pixel colour corresponding to the image projected onto this default layer is used.

3.3 DSI Smoothness

The simplest cost function evaluated for determining agreement between different camera views considers only the maximum absolute difference in intensity values for the pixels in the overlapping regions on the cylindrical projections. For example, if frames from three different cameras overlap at a particular pixel on

a cylinder with 8-bit greyscale intensity values of 140, 150, and 170 respectively, the disagreement value for this location would be 20. As it is standard practice to represent perfect agreement as white, the disagreement value is subtracted from the maximum intensity value of 255 to give the final colour encoding of 235 for this example. This single pixel based approach introduces graininess in the stitched images which did not exist in the source images.

Experimentation with cost functions which evaluated absolute difference over variously sized windows saw decreases in grain. For any given pixel in an overlapping region a window is centred on that pixel. Each of the pixels in the window are given a disagreement value which, unlike the example above, are the average absolute disagreement between projections instead of the absolute maximum disagreement. These disagreement values are averaged across the entire window and the pixel surrounded by the window is assigned the corresponding colour encoding. Windows with dimensions varying from 3×3 up to 19×19 were tested with no noticeable continued improvement beyond window sizes of 11×11 for source images with 1280×1024 resolution. This reduction in grain can be observed in Fig. 5 for windows of dimensions 5×5 to 11×11 . Let

$$C_{DSI} = \frac{\sum_{i=0}^w \sum_{j=0}^h [|C_{ij}^0 - C_{ij}^1| + |C_{ij}^0 - C_{ij}^2| + |C_{ij}^1 - C_{ij}^2|]}{3ij} \quad (5)$$

This describes a cost function for a window of arbitrary size and in an overlapping position covered by three cameras. C_{DSI} is the colour for a pixel in a DSI. It is determined by a window with width w and height h . In C_{ij}^k , k is the index of the camera.

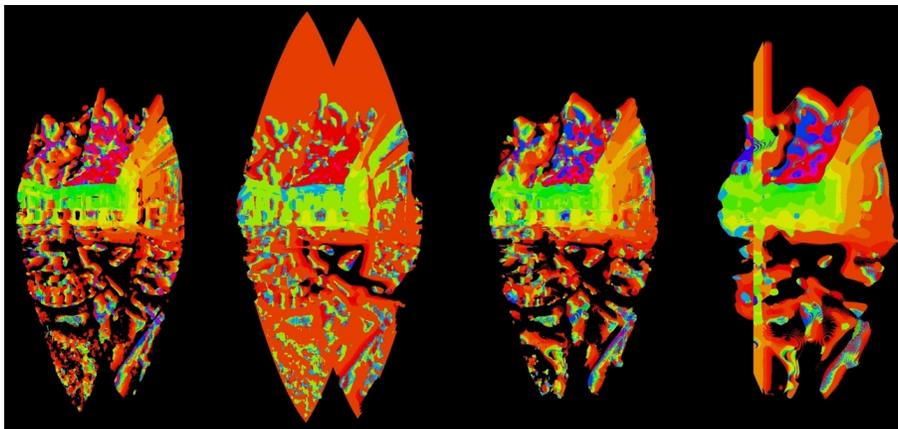


Fig. 4. From left to right: the original DSI set as produced with an 11×11 cost function window, a Gaussian blur, median blur, and simple blur, all with 9×9 windows.

In order to evaluate different approaches to creating DSI sets it was necessary to visualize entire DSI sets at once. To create a single image showing an entire DSI set the same method as was used in negotiating pixel values for the overlapping area was used. Instead of using pixel colours as determined by the cameras, colours were instead assigned according to the radius of the cylinder from which they came. Pixels whose highest level of agreement was determined to have come from the minimum radius cylinder were colored red, while pixels which were found to be in agreement at the maximum cylinder radius were colored violet. Cylinders in between were given corresponding values from the intervening spectrum. Pixels which reported extremely high or low levels of agreement were coloured black. The result approximates a coloured depth map and allows for comparison of different DSI treatments.

To further reduce the noisiness of the DSIs Gaussian, median and simple blurs from the OpenCV library were employed with window dimensions ranging from 3×3 to 15×15 . The best results were obtained using a window with 9×9 dimensions and a simple blur. The resultant smoothing of the DSI set for each of the blurring techniques can be seen in Fig. 4

3.4 Bird's-eye View Distortion

The final step in producing the desired illusion of an image produced by a camera floating a few meters above a car and facing its roof, is the distortion of the cylindrical image produced in the stitching process described above. In order to do this a 3D bowl-shaped object was produced and the cylindrical image was mapped to this surface as illustrated in Fig. 5. By producing a view which looks into this bowl, it is possible to produce the desired effect.

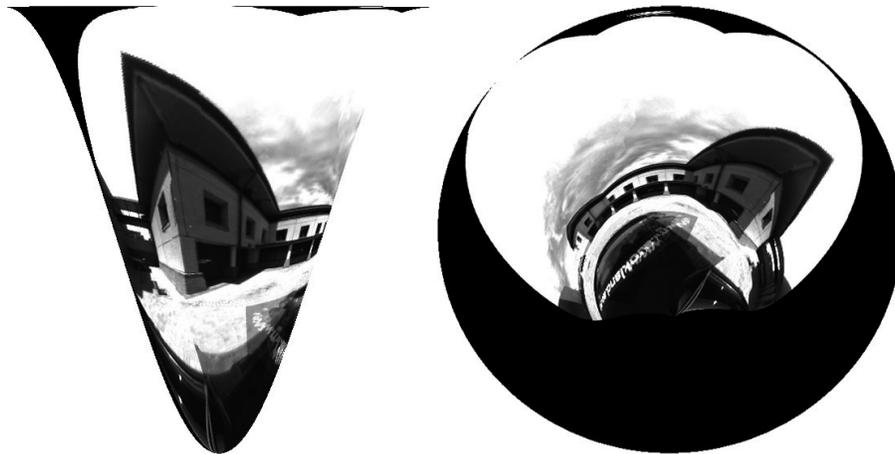


Fig. 5. *Left:* Side-on view of birds-eye projection surface. *Right:* Top-down view into projection surface.

4 Conclusion

The cylinder sweep approach to stitching wide angle images into omnidirectional cylindrical images can produce images useful for driver assistance systems. The bird's-eye view perspective in particular is helpful as it combines an omnidirectional perspective to drivers which is also intuitively interpretable.

The stitching process does not produce visible seams within areas of image overlap, but the transition between areas covered by multiple cameras and areas covered by single cameras can sometimes produce unsightly seams. The use of a default layer is not a suitable solution to this issue.

The cylinder sweep process could also be improved by developing an approach that does not rely on a finite number of cylinders at discrete distances, but allows for the definition of a continuous function for image projection comparison.

Acknowledgement: The authors acknowledge the support by Je Ahn in the experimental studies.

References

- [Bouget 2010] Bouget, J.-Y. (2010). Camera calibration toolbox for Matlab. Retrieved January 28, 2011, from http://www.vision.caltech.edu/bougetj/calib_doc/.
- [Collins 1996] Collins, R. (1996). A space-sweep approach to true multi-image matching. In Proc *CVPR*, pages 358–363.
- [Gehrig et al. 2008] Gehrig, S., Rabe, C., and Krueger, L. (2008). 6D vision goes fisheye for intersection assistance. In Proc. *Canadian Conf. Computer Robot Vision*, pages 34-41.
- [Kang et al. 2004] Kang, S. B., Szeliski, R., and Uyttendaele, M. (2004). Seamless stitching using multi-perspective plane sweep. Microsoft Research, Technical report MSR-TR-2004-48.
- [Liu et al. 2008] Liu, Y.-C., Lin, K.-Y., and Chen, Y.-S. (2008). Bird's-eye view vision system for vehicle surrounding monitoring. In Proc. *Robot Vision*, LNCS 4931, pages 207–218.
- [Rodrigues 1840] Rodrigues, O. (1840). Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *J. Mathématiques* **5**:380–440.