# Relative Disparity Using Motion in Stereo Vision

G. Searle, R. Green

Department of Computer Science, University of Canterbury, Christchurch, New Zealand

Email: gws40@student.canterbury.ac.nz, richard.green@canterbury.ac.nz

## Abstract

In this paper we propose a method which uses motion in the field of view of two un-calibrated cameras to produce a disparity map relative to the cameras. Current stereo vision initialisation techniques focus on time consuming manual calibration of stereo cameras. This research demonstrates the use of a novel algorithm which uses motion to determine conjugate pairs of points. The disparity map is automatically generated utilising optical flow of objects for motion common to both cameras. A basic form of reconstruction is then used, without the intrinsic and extrinsic parameters, to determine a relative disparity. This enables a "plug and play" implementation of stereo vision.

**Keywords**: Stereo vision, camera calibration, conjugate pairs, motion, relative disparity

## 1 Introduction

The goal of this paper is to use uncalibrated cameras with stereo vision system to avoid the traditional time consuming manual process of camera calibration. Currently techniques in stereo vision usually involve an initial calibration period which attempts to acquire the global surface information. This information calculates both cameras' field of view, focal length and spatial relationship to each other. This introduces a method to calculate approximate relative disparity without the need for such information.

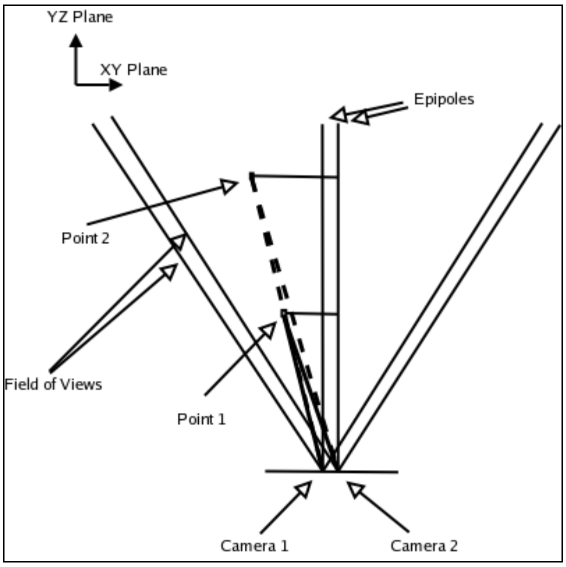


**Figure 1:** The setup of the cameras with two points which would occur as the same pixel in Camera 1. The dashed lines represent the views of Point 2.

Stereo vision uses epipolar geometry such that two convergent cameras' views intersect at a single point. Parallel epipoles (Figure 1) are difficult to accomplish in practice [1]. However two matched cameras have vertical symmetry allowing this alignment to be less problematic in the approach taken by this research. There is minimal difference in the final result if the cameras are slightly convergent or divergent.

**Figure 2:** A single frame from footage with both of the extremities shown with white vertical lines. This example has a slight epipole convergence.

Motion has been used to determine the conjugate pairs where such conjugate pairs were taken in a one dimensional approach. The absolute difference of motion between the two frames is first calculated. The extremities of the horizontal axes are then taken to be pairs of the cameras' respective horizontal extremities. This can be seen in Figure 2. This assumes that if both cameras are in the same plane and in the same direction that the movement seen in one camera is seen in the other. The accuracy over distance is proportional to how far apart the cameras are.

It was found that the one dimensional accuracy in this approach was much more accurate than the two dimensional approach for two reasons. The first is that two dimensions require accurate cameras alignment so they are approximately pointing in both the same XZ plane and YZ plane. The second is because most images with movement would intersect one of the boundaries for the camera most of the time. This boundary intersection results in a false read for the camera.

## 2 Background

A large amount of research has gone into stereo matching, camera calibration and fast algorithms for stereo vision [13]. This paper does not look at definitive matching or camera calibration.

Stereo vision is not commonly used principally because of the correspondence problem [3]. The implementation described in this paper provides a novel way to avoid the correspondence problem.

The papers which consider automatic camera calibration are constrained to particular environments and therefore cannot be generalised. One example of this is that one paper [3] calibrated the cameras' focal point by using a robotic rotation.

Research that has involved uncalibrated stereo vision often involves a fiducial marker [14].

The significant limitations of stereo vision in prior research, are the correspondence problem and the reconstruction problem [1] which have been overcome in this implementation. However this proposed approach also creates new limitations caused by the unknown intrinsic and extrinsic parameters.

## 3 Relative Distance

Because there is no accurate information about the focal length of the cameras or the distance that separates the cameras, all calculations are relative. In this approach, the absolute distance is not calculated because the absolute distance requires additional calibration to relate them to the image-derived data. Instead the relative distance is calculated, which has been said to be useful [7],[12].

In this paper the centre of the camera image is assumed to be the optical axis. The left-most and right-most pixels are assumed to be equal angles from the optical axis.

To find the relative disparity:

1. Calculate the difference of motion on each camera.
2. Gaussian filter the result.
3. Find the x, y limits for each camera.
4. Calculate the distance of each pixel from the centre point of the camera.
5. Compare the results with each camera.

Figure 1 and Figure 3 both show that when a point in motion is further away from the camera it will have a smaller disparity than a point closer to the camera where the ratio between the disparities is closer to one.

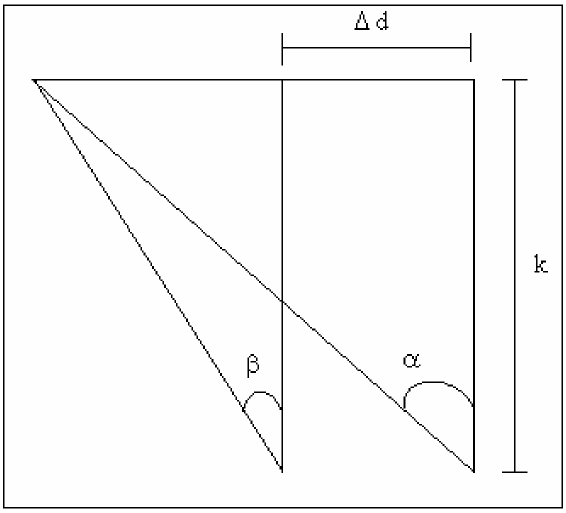


**Figure 3:** A diagram showing the setup of the cameras. Δd is the extrinsic distance between the two cameras.

Figure 3 shows the mathematical and computational efficiency of assuming both epipoles are parallel. By assuming this we can calculate the relative distance k by:

$$k = \frac{\Delta d}{\tan(\alpha) - \tan(\beta)}$$

α and β would be calculated using the intrinsic parameters, these are not available. Δd is the distance between cameras. This is an extrinsic parameter that is also unknown. Since both cameras are assumed to have a field of view of less than 180° one finds:

$$\lim_{(\alpha - \beta) \to 0} \frac{\Delta d}{\tan(\alpha) - \tan(\beta)} = \infty$$

When the motion is at an increasing distance from the cameras, the angles between the pixels on relative cameras will have a smaller difference between them. By taking the ratio between these pixels, relative distance can be produced.

A constraint inherent in this implementation is the assumption that both the cameras have the same focal length as is generally the case when acquiring two identical cameras at the same time.

## 4 Experimental

There were three different experiments taken. The first measured the relative distance of a person moving backwards and forwards in front of the camera. For the second experiment, a person stood in the middle of the screen and rotated their shoulders. For the third, relative disparity was calculated for movements at a measured distance.

The first experiment involved a person walking backwards and forwards in front of the camera. The person moved at an approximately constant speed. They moved between 70 to 170 cm from the cameras. While moving the person also waved their arms around. There are two reasons for this:

1. To make sure that every frame has registered movement. The movement is calculated by the absolute
2. difference between two frames: if no movement was made, there would be a false read.
3. If a single camera measured a person walking backwards a forwards it could tell if a person was moving towards or away from the camera due to the size change. Making the person move their arms around this does not allow the size of the object to measure the relative disparity.

Experiment two had a person stand at 1 metre from the cameras and rotated their shoulders. Both the left most point and the right most points were taken. The person did not rotate their shoulders to 90º. This is because at 90º one shoulder is obscured from the camera and its distance cannot be calculated. The relative disparity was calculated by the average of the difference between the two cameras respective points. Thus a high relative disparity indicates a point close to the camera. It was predicted that the there would be two harmonic waves out phase. When one point is close to the camera, the other point is far away.

Experiment three measured relative disparity at known distances. At a measured distance movement was made and the left-most points from both cameras were taken. Only the left-most points were taken because the movement was made from the right side of the camera out of the field of view and could not be measured. Over a period of 80 frames the average relative disparity value was calculated. The relative disparity is the absolute distance in pixels between the left-most points on each camera. This means that a point close to the camera will have a high relative disparity. The disparities were taken at 20 cm intervals.

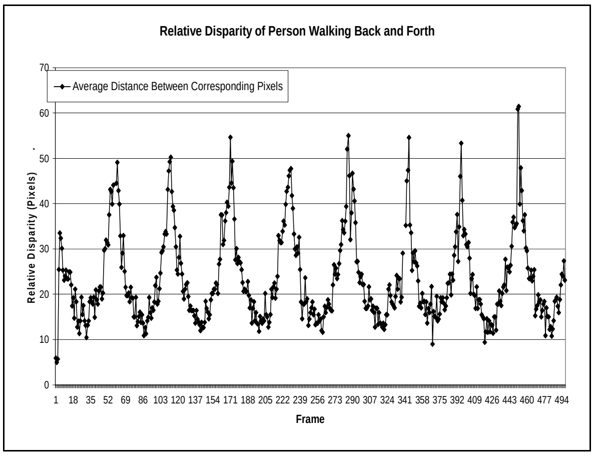


**Figure 4** Results of Experiment 1. Showing the relative disparity of person moving backwards and forwards in front of a camera.

The relative disparity is calculated by the average distance between pixels. Thus a higher value corresponds to a distance that is close to the camera

## 5 Results

Locations and rotations were successfully calculated in real-time using relative disparity from motion with the following computer vision platform:

1. Intel Pentium 4 2.8GHz Processor
2. 1 GB PC3200 DDR400 RAM
3. XP Professional SP2
4. GeForce FX 5200 128MB

Both of the cameras used were USB Logitech Camera Pro 4000. The cameras were operating at 640x480 resolutions and 30 fps at ambient indoor illumination levels.

Figure 4 shows the results of Experiment one. Experiment one involved measuring the relative disparity of a person moving backwards and forwards in front of the cameras. It is seen on Figure 4 that the curve is harmonic. This was predicted because the person is moving backward and forwards at a constant speed, and has to turn around.

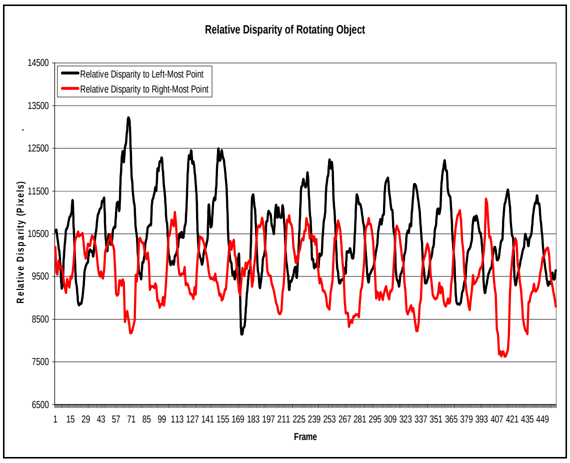


**Figure 5:** Results of Experiment Two. Showing the relative disparity of a person rotating their shoulders

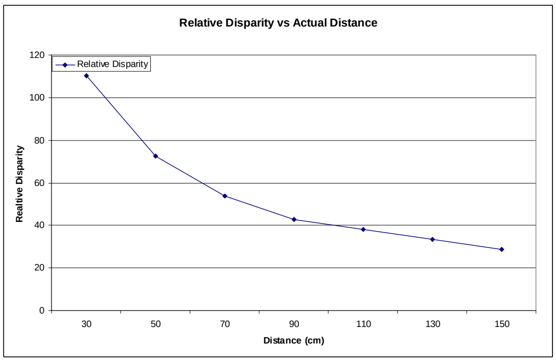


**Figure 6:** Results from experiment three. Showing the relative disparity against the measured distance.

Figure 5 shows the results of the second experiment. The second experiment involved a person rotating their shoulders at a fixed distance. It is seen in Figure 5 that the phases of each side are out of phase. This was the expected result. The left-most point values were lower than the right's on average. Possible explanations for this are that there was a slight convergence in the cameras' alignment or the absolute difference was sharper for the right side than the left.

Figure 6 shows the results of experiment three. Experiment calculated the relative disparity at a measured distance. The expected results would be to have a linear line. The actual results show a much higher relative disparity when close to the camera. However as the distance gets farther from the camera, the relative disparity gets closer to linear.

## 6 Conclusion and Future Work

It is possible to have "plug and play" stereo vision. Using this algorithm we have been able to recognise relative differences and rotations of objects between distances of one to five meters for two cameras positioned about 8cm apart.

This implementation demonstrated a computationally efficient algorithm for stereo vision containing motion without the need for time-consuming stereo camera calibrations. The extreme points of motion can be used as good conjugate pairs.

All three experiments were more successful than predicted. Every expected result was reached. This proves that plug and play stereo vision is very possible.

Future research will investigate dynamic adjustment of the Gaussian filter to lower the errors associated with motion at very near and far distances and for variable camera separations.

Using a conjugate pair detection algorithm [4],[5],[11], instead of taking only two points, will be further investigated to determine a more accurate disparity. As the conjugate pairs have proven to be successful, it could be possible to track the optical flow of all points within this region.

## References


[1] E. Trucco, A. Verri, "Introductory Techniques for 3-D Computer Vision", New Jersey: Prentice Hall, 1998

[2] U.R. Dhond, J.K. Aggarwal, "Structure from Stereo – A Review", IEEE Trans. Systems, Man, and Cybernetics, 19(6) 1989, pp 1489–1510

[3] P. Ho, R. Chung, "Stereo-Motion with Stereo and Motion in Complement", Pattern Analysis and Machine Intelligence, IEEE Transactions on; 22(2), 2000, pp 215–220

[4] G.A. Jones, "Constraint, Optimization and Hierarchy: Reviewing Stereoscopic Correspondence of Complex Features", Computer Vision and Image Understanding, 65(1), 1997, pp 57–78

[5] T Shen, C Menq, "Automatic camera calibration for a multiple-sensor integrated coordinate measurement system", Robotics and Automation, IEEE Transactions on; 17(4), 2001, pp 502–507

[6] S. Shih, Y. Hung, W. Lin, "Accurate linear technique for camera calibration considering lens distortion by solving an eigenvalue problem", Optical Engineering 32(01), 138-149, Brian J. Thompson; Ed. 1993

[7] J. Aloimonos, I. Weiss, A. Bandyopadhyay, "Active Vision" International Journal of Computer Vision; 4(1), 1988, pp 333–356

[8] R.C. Gonzalez et al., “Fast stereo vision algorithm for robotic applications”, Emerging Technologies and Factory Automations, 18-21 October 1999, pp 97–104.

[9] E. Grosso, M. Tistarelli, “Active/Dynamic Stereo Vision”, Pattern Analysis and Machine Intelligence, IEEE Transactions on; 17(9), 1995, pp 868-879

[10] H. Sunyoto et al, “A Comparative Study of Fast Dense Stereo Vision Algorithms”, 2004 IEEE Intelligent Vehicles Symposium, 14-17 June 2004, pp 319–324.

[11] D Yuan, R Chung, “Correspondence-free stereo vision for the case of arbitrarily-positioned cameras”, Robotics and Automation 2003, 14-19 September 2003, pp 1688–1693.

[12] G. Sandini, M Tistarelli, “Active tracking strategy for monocular depth interference over multiple frames”, Pattern Analysis and Machine Intelligence, IEEE Transactions on; 12(1), 1990, pp 13–27

[13] R. Kellly, P. McConnell, S Mildenberger, “The Gestalt photomapping system”, Photogrammetric Engineering and Remote Sensing, 43(11), 1977, 1407–1417 [10]

[14] Y. Kuno et al., “Human-Robot Interface Using Uncalibrated Stereo Vision”, Intelligent Robots and Systems 95, IEEE/RSJ International Conference on; 1, 1995, pp 525–530