Tracking Object Trajectories Relative to Planar Surfaces Using Stereo

November 7, 2007

Matthew Elliot

Supervisor: Dr. Richard Green

Department of Computer Science & Software Engineering

University of Canterbury, Christchurch, New Zealand

Abstract

This project proposes a methodology for 3D tracking of objects in relation to a planar surface, with trajectory accuracy enhanced using applied statistical analysis. Planar surface extraction, with camera position and orientation invariance, is achieved by finding limiting regions established by graph-based segmentation and mapping the resulting segments to disparity data from a stereo camera. Secondly, object detection and tracking is performed using a combination of adaptive background subtraction and least squares linear regression for calculating object trajectories. The accuracy of bounding planar surface extraction is shown to be accurate to within 1.4% and tracking has shown similar high correlations between the calculated and actual positions.

Contents

List of Figures

1 Introduction

This paper proposes a new methodology for tracking objects and their associated trajectories using a novel bounding planar surface detection technique and applied statistical analysis to extract trajectory information from disparity data. Tracking can be considered a two part problem; finding the object itself and determining its position in relation to the scene. The first part of this project focuses on establishing a reference plane for tracking by determining the location of the ground plane from stereo images. This process uses graph-based image segmentation to separate the image into regions. The average position of each of these regions is established, and inference about the relative positions and ultimately extraction of a planar representation is made. This process also uses standard matrix rotations to normalise disparity data points back to a standard reference frame, allowing extraction of bounding planes with camera rotation invariance. The second part of this project focuses on object detection, tracking and trajectory calculation. Object detection is accomplished using adaptive background subtraction. The resulting difference image is then mapped onto the disparity data from the stereo camera and positional information about the object is calculated. If a series of observations are made (greater than five), then linear least squares regression is used to find the plane in which the object's trajectory lies, and each recorded position is translated onto that line. Finally, velocities and positional information relative to the ground plane are calculated. Successful results with high accuracies for ground planes and moderate accuracies for trajectories suggest that the proposed method would be useful for tracking objects in 3D space.

1.1 Extraction of Planar Surfaces

Limitations on the view created by bounding planar surfaces are useful because they provide a limit on the distance objects can appear in a particular direction. For example, anything beyond an internal wall is obscured from view. Being aware of this constraint allows us to conclude that if there is such a bounding planar surface it will occur at this limit. A simple technique that would take advantage of this would be to assume the bounding planar surface lies at the limiting value and consider it solved. However, this approach is unlikely to provide the accurate location of planes as the disparity data from a stereo camera is noisy with discontinuities arising from lack of texture and poor lighting conditions. These issues lead to the need for determination of the group of 3D points that make up the bounding planar surface. Figure 1 is the disparity map from a simple scene.

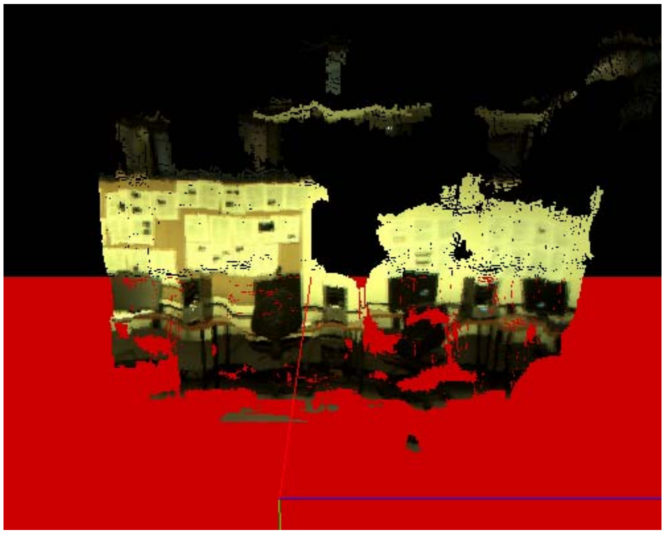


Figure 1: Disparity data from a simple scene with the calculated ground plane superimposed as a red region.

Grouping similar points into homogeneous regions in images is encompassed by image segmentation. Colour image segmentation is a difficult problem with substantial research focusing on this area. W. Skarbek and A. Koschan [1] provide a broad summary of techniques and their limitations. For image segments to accurately represent contiguous regions in the scene, the technique for image segmentation must tolerate small colour variations. Lighting effects on grass (shadows) or small texture variations in carpet colours or patterns may cause such problems. A technique that is adaptable and copes well with this is graph-based image segmentation [2].

To determine the plane's parameters, graph-based image segmentation is used to break up a scene into a subset of regions. By individually mapping each of these regions onto disparity data from a stereo camera, the region representing the ground can be found. Figure 1 shows the scene after the bounding planar extraction process has occurred, targeting only the floor.

This technique of bounding planar surface extraction is evaluated and the results are very promising with accuracies all falling within with 6.7% percent from the true plane for single frames and 1.4% from the true position of the plane when considered across multiple frames (running average). This accuracy could prove useful with future work where ground plane extraction is key, as in robot navigation for example.

1.2 Object Detection and Tracking

Object detection and tracking are both interesting problems and there are many possible solutions to both. Object detection is about determining whether an object appears in a image or not. For this project, the objects of interest are those that are moving and are characterised by changes in the image. Based on this fact, it is possible to extract from consecutive images an object's location. This project considers a static camera setup which enables a *stationary camera, moving object* (SCMO) [3], detection technique to be applied; in this case adaptive background subtraction. Once the pixels representing an object are located, they can be used in conjunction with the disparity data from the stereo camera to determine the 3D location of the detected object.

Object tracking entails detecting an object in multiple consecutive frames; possibly predicting where it might appear in the next frame, (Kalman Filter) [19]. In this project, tracking is attempted using known physics properties and applied statistical analysis is used to estimate object trajectories. The main assumption made here is that an object follows a simple linear ballistic trajectory, which allows linear least squares regression to be used to estimate its path. The recorded positions of the object of interest are then mapped to this line of best fit giving an adjusted representation of the recorded trajectory. Evaluation of the positional accuracy of these adjusted points is then considered. Single observations have proven to be of varying accuracy due to noise. Averaged observations (ten observations), have shown a high correlation with true positions and reflect what may be possible with higher camera resolutions, improved light levels and more effective disparity algorithms. Future work may also involve modification of disparity algorithms to better suit detection and tracking of a single object.

2 Background & Related Work

As established in the introduction there are two main proposals in this research, planar surface detection and object recognition and tracking using trajectory based analysis. Also of relevance are the concepts and processes of calculating disparity maps as this project uses solely stereo vision for 3D positional data acquisition. This section outlines these three areas and important related work relevant to this project.

2.1 Stereopsis

To reconstruct a scene and find bounding planar surfaces, the need to extract accurate 3D information becomes apparent. There are a couple of standard techniques to extract this information, including stereopsis (multiple cameras) and consecutive images taken from a single, moving camera [3]. Using either technique, the following overview of stereo can hold but it is focused more at stereopsis.

2.1.1 Stereopsis and Processing

There are many different methods for calculating disparity but they are broadly classified into three distinct categories; feature-based, area-based and phase-based methods. D. Scharstein and R. Szeliski [4] examine a wide range of stereo algorithms and outline some of the tradeoffs present in common approaches resulting from occlusions, lighting effects and texture density.

2.1.2 Disparity

To calculate disparities, the geometry of the stereo system must be known. Binocular stereo in its simplest form requires that two views be separated only in the x-direction by a known distance (baseline). Any feature in a scene will occupy a different position in each view, with the displacement between these two positions determining the disparity. Figure 2 shows a simple binocular system. There are two important aspects to this system, the epipolar plane and epipolar line. The epipolar plane is defined as the plane passing through the two view centres (C_1, C_2) and the point of interest in the scene (P). The epipolar line is determined by the intersection of the epipolar plane with view plane. The difference in distance when the views are superimposed determines the disparity of a point. The relationship between disparity and depth is related and can be calculated using triangulation.

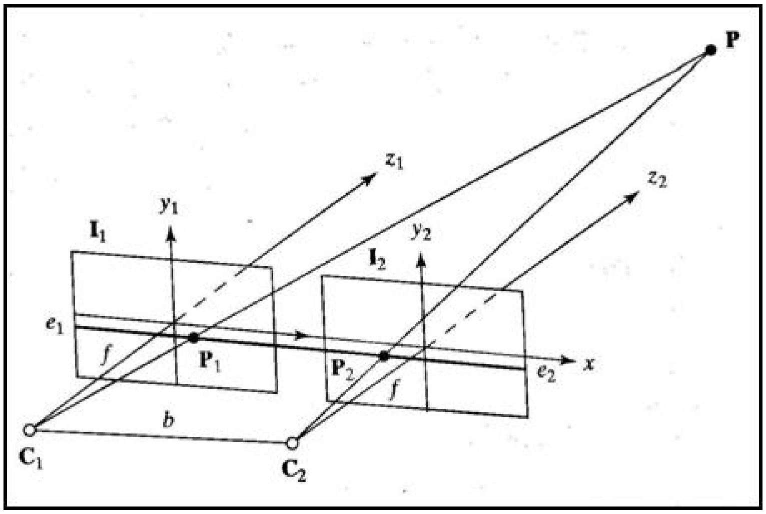


Figure 2. The process for calculating depth from two stereo images

When calculating disparity the largest influence is the horoptor of the stereo system. The horoptor is a region in space that maps exactly to the two images from the cameras [5]. In an active stereo camera system, this region can be made specific and disparities calculated based on that region. This essentially allows focusing of the stereo camera disparity calculations at a specific depth. Formally disparity is found by the following equation:

$$z = \frac{bf}{(x_L^{'} - x_R^{'})} \tag{1}$$

Where z is the distance to the point in the scene, b is the baseline length, f the focal length, $x_L^{'}$ the displacement of the x position of the point in the left view and $x_R^{'}$ the displacement of the x position of the point in the right view.

2.1.3 Drawbacks and Limitations

Stereopsis itself is a non-robust technique for extracting depth information, and knowledge in advance of its limitations is critical to application development. Stereopsis limitations are caused by insufficient colour contrast, low or uneven lighting, lack of texture, and occlusions. Colouring becomes an issue in stereo vision when a foreground object has the same colour to that of a background object. Stereo vision algorithms rely on correlation between stereo images and colour contrasted edges play a large role in this process. The depth data for such an area is highly likely to be incorrect. Lighting can also cause problems like colour, especially specular effects, where the light intensity varies greatly making correlation impossible. Large, untextured regions can also prove troublesome because correlation can often rely not just on colour but image features. This is clearly going to cause further problems with correlation and hence depth calculations. Partial occlusions are also a significant problem, where if a point is not visible to both cameras than it is never possible to calculate disparity for that point.

2.2 Planar Surface Detection

Once the extraction of accurate 3D information is achieved the focus lies on techniques for scene reconstruction; here the task of planar surface extraction is closely examined. Extracting planar surfaces has been solved with both mono and stereo camera systems but there is limited work on plane extraction using stereo. This section will explore stereo processing and review existing techniques for mono and stereo planar detection, outlining advantages and limitations for each.

2.2.1 Ground Segmentation

B. Liang and N. Pears [6, 7] developed a technique that focuses primarily on ground segmentation from images for a mobile robot application using a mono camera system. Using consecutive images for distance calculation and colour, contours, corners and their motion, they successfully extracted the ground plane region from the images and further used it for robot navigation. While successful, the accuracy of the data from consecutive mono camera images is likely to be less accurate than a calibrated stereo system. This means a stereo system should be able to extract more accurate depth information. Furthermore, their technique is only used in locating the planar region representing the ground in the image and does not deduce the mathematical equation of a plane that represents that surface. Figure 3 shows the result of their ground plane segmentation.

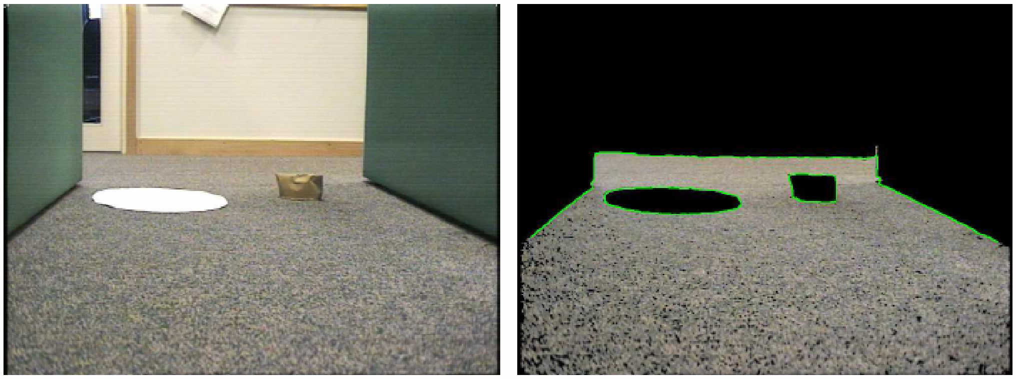


Figure 3. Image from B. Liang’s and N. Pear's [6, 7] papers illustrating their plane segmentation technique. The plane that is produced is a visual aid and no plane parameters are calculated.

2.2.2 The 3D Hough Transform

The 3D Hough Transform arose from a simpler 2D technique for extraction of lines. The original Hough Transform methodology was created by Duda and Hart [8] and extended and popularised in computer vision by D. Ballard [9]. An example includes B. Yu and A. Jain’s [10] detection of road markings. The 3D Hough Transform has become a predominant technique for extracting planar surfaces and is described as follows. A plane (*P*) can be expressed as follows in 3D space:

$$P = x\sin\theta\cos\phi + y\sin\theta\sin\phi + z\cos\theta \qquad (2)$$

Where (p,θ,ϕ) is a vector, perpendicular to the plane originating from the origin (*O*). 3D Hough space is also defined by the parameters *P, θ,* and ϕ which can be mapped to a standard 3D space (*x*, *y*, *z*). For any point (x_i, y_i, z_i) in 3D space there will be many planes that pass through the point. The planes are described by the equation:

$$p = x_i\sin\theta\cos\phi + y_i\sin\theta\sin\phi + z_i\cos\theta \qquad (3)$$

This equation will now express all planes that pass through the point (x_i, y_i, z_i) in terms of the space defined by *P, θ,* and ϕ. It follows that in 3D Hough space there will be a curved surface representing each point from the standard 3D space. Points that fall on a plane in standard 3D space will all share the commonality that their curved surface in Hough space will intersect with the point (p,θ,ϕ).

To find such a plane, each point in the standard 3D space must vote for a plane representation in Hough space. A 3D histogram is established in the space defined by *P, θ,* and ϕ and for each point, (x_i, y_i z_i), the bins for which that point’s curved surface in Hough space crosses are incremented. After all points have been considered, the bin with the largest count is used to determine the plane parameters.

An example of this process in practice was H. Wu, G. Yoshikawa, T. Shioyama, T. Lao and T. Kawade’s [11] paper where they used a 3D Hough Transform to detect the plane in which the frame for a pair of glasses appeared. This allowed them to separate the glasses from a face when performing face recognition, greatly enhancing the accuracy. The 3D Hough Transform is an effective way to extract planes from 3D data but it suffers a large limitation in the time it takes to calculate plane parameters.

2.3 Object Detection and Tracking

Once a planar surface is established as a reference for tracking, the second part of the overall problem can be approached. This section breaks this problem down into a further two sub-problems; recognition of an object, followed by tracking and calculation of the trajectory of the object in 3D space.

2.3.1 Moving Object Detection

The substantial research on detection of moving objects in images can be split into two broad categories; *stationary camera, moving objects* (SCMO) and *moving camera, stationary objects* (MCSO) [3] Stationary camera techniques rely on the camera being static. This reduces computational complexity, as most of the scene will remain unchanged between frames. Some stationary techniques include, image differencing (background subtraction, frame differencing, double frame differencing). Moving camera techniques can generally cope with both static and non-static circumstances because they track features of the scene itself. Examples of this technique include optical flow and template matching. These techniques all have varying benefits and hence tradeoffs are evident between them.

In its simplest form, background subtraction is performed by subtracting the current image from that of a reference frame image. That is for every pixel in an image, subtract the value of the corresponding pixel in the reference frame to determine if there is any change at that pixel. S. Kamath [12] gives a good description of varying background subtraction techniques in his paper on tracking traffic. The paper also gives an overview of different recursive and non-recursive background subtraction techniques and compares the accuracy under a number of different conditions.

Recursive techniques require that a single background model be updated after each frame. This requires little storage but can prove less adaptive as erroneous frames can affect the background model over a long period. An example of a recursive technique is the Kalman Filter [13]. The Kalman Filter is used to predict the unknown state (*S*) of a pixel at a particular point in time (*T*). The estimate of the pixel's state is based on a measurement (*M*) at a given time. The Kalman Filter also estimates the uncertainty of the given pixel state estimate. Recursive techniques, while often providing high accuracies, do so at the cost of performance as the algorithms are more computationally expensive.

Non-recursive techniques on the other hand keep a record of the last *N* frames. This is good because they can adapt rapidly as the background model changes but suffer from the amount of storage space they require. Probably the simplest and most common non-recursive technique is single frame differencing. Single frame differencing is performed by subtracting the current frame C^F from the previous frame $C^F - 1$. P. Rosin & T. Ellis [14] examine in depth the process of image differencing and propose strategies for thresholding. A correctly set threshold can both eliminate noise and prove useful in dealing with slow ambient light changes in a scene. Unfortunately, there are other significant weaknesses in single frame differencing including the aperture problem (holes in an object caused by slow movement) and ghosting (two objects appearing; one where the moving object is, one where the moving object was).

2.3.2 Object Tracking Using Stereo Vision Systems

Object tracking using stereo vision systems is an interesting area of research because the problem is inherently difficult. There are a number issues that arise which research in the area has attempted to alleviate. S. Rougeaux, N. Kita, Y. Kuniyoshi, and S. Sakane's [15] paper was an early paper that aimed to track objects against complex background scenes using a stereo camera system. They propose a method that allows automatic updating of the horoptor during the tracking process based on what they call 'virtual horoptors'. This paper is based upon work on dynamic vision systems [15-17] where the cameras can track similarly to the way the human eyes do; following an object. This requires a complex mechanical device for manipulating camera directions and is an expensive area.

Further research that uses this same premise is M. Tanaka, N. Maru; and F. Miyazaki's [16] paper where a robotic, binocular vision system that can simulate human eye movement is used for 3D tracking of objects. The unique part of their system is their localisation of the object using disparity-based segmentation. They further describe this technique as a filter that separates disparities of the object of interest from distracting background disparities. By using their proposed disparity-based segmentation technique they show that fixation on the object's surface is possible, increasing object tracking accuracy.

Other research with stereo tracking systems has focused on tracking more complex objects such as people [18, 19], people's heads [20] and traffic [21]. Tracking of more complex objects is obviously more difficult and hence more elaborate measures (hybrid measures) are employed in these papers. R. Muñoz-Salinas, E. Aguirre and M. García-Silvente [19] use a combination of face detection and colour to keep track of people while using the stereo camera to give 3D localisations for each. They also use a Kalman Filter for predicting subsequent positions of each person being tracked. Unfortunately, using a Kalman Filter for human tracking is not robust because people can change direction quickly, causing a Kalman Filter based system to fail. N. Setiawan, S. Hong, and C. Lee [18] proposed a different approach using fragment-based histogram matching. For each person in the scene a representation in the form of a fragment map is established and determination of a particular person is achieved by a voting system. Their work has shown good initial results. D. Russakoff and M. Herman's [20] paper focused on a simpler object, tracking of only a persons head. Their research is targeted at the area of smart environments and perceptual user interfaces, where body pose information needs to be accurately extracted. They focus on background/foreground segmentation for object localisation and use edge detectors on the foreground to find occluding edges. From their research, they have shown that a person's head and pose can be extracted accurately and tracked through 3D space. Finally, Y. Wakabayashi and M. Aoki [21] use a stereo system for tracking traffic; more specifically traffic flow. Their results were generally positive and even allowed vehicle classification to some extent. The benefits of this system included road surface and shadow invariance.

This prior research suggests that stereo cameras are an effective tool in tracking objects in 3D, but with significant drawbacks. Some of these issues can be counteracted by using hybrid techniques, which incorporate other standard computer vision algorithms, but accurate disparity data coupled with varying user interpretation still makes object tracking using stereo vision systems overly difficult. None of these papers attempt object trajectory calculation to smooth results from the tracking process.

3 Design and Implementation

3.1 Overview

The design and implementation of the methodology presented here is broken down into three distinct parts; finding the location of the ground plane, object detection, and tracking. The location of a ground plane is established using the new method proposed in this paper. In the simplest case, it involves, segmenting the image, matching each segment to the depth information, and finally finding the plane parameters. To detect objects, a differencing algorithm is used to find movement in the images, the difference image is then passed to an object search algorithm that classifies close groupings of difference pixels as objects, and finally object elimination is performed based on a couple of heuristics. Object tracking requires that the same object be located in near sequential frames. Trajectory information is extrapolated by examining ball positions in consecutive frames.

3.1.1 Small Vision System

The Small Vision System (SVS) software package is used to interface to all Videre stereo camera systems. It provides functionality to set up the camera and control image acquisition. It also provides configurable, accurate disparity calculations from a set of stereo images. This allows the 3D position of every pixel with calculable disparity to be determined. The SVS package also provides C/C++ and Matlab libraries and programming APIs.

3.2 Planar Surface Extraction

In determining bounding planar locations, a new method was developed. Most existing techniques for finding planar surfaces in images use some fitting process like the 3D Hough Transform [11, 22] to extract planes from the 3D data gained from stereo processing. The technique proposed here works in the reverse way by mapping image regions, gained using image segmentation, onto the 3D data and then uses these regions to infer information about the scene. This section describes the process of extracting bounding planar surfaces using this technique and knowledge about the scene.

The extraction of bounding planar surfaces can be broken down into a series of simple steps including.

- Scene deconstruction using graph-based image segmentation.
- Region to 3D position mapping.
- Scene rectification.
- Determination of plane parameters.

The first three steps in this process consider the scene as a whole and form the basis for the extraction of any bounding plane in step four. The following sections further detail the process for bounding planar surface extraction.

3.2.1 Image Segmentation

The first step in determining where a bounding planar surface is located is to segment the image. This implementation uses P. Felzenszwalb and D. Huttenlocher’s graph-based image segmentation algorithm [2]. This algorithm works by trying to create an undirected graph:

$$G = (V, E) \tag{4}$$

Where $v_i \in V$ are the set of pixels from the image to be segmented, $(v_i, v_j) \in E$ gives the correspondence of two neighbouring pixels, and $\mathrm{W}(\mathrm{v}_\mathrm{i}, v_j)$ gives a measure of correlation between two pixels and hence determine their associativity for grouping into regions.

After the algorithm has run, the connected regions of the graph will determine each distinct segment. The benefit of this technique is that the weight function can be adjusted so that slow or subtle variations of colour caused by lighting (shadows) or minor texture can be considered a single segment. The left colour image from the stereo pair is segmented in this way with the output being a segmented image with each region being represented by a different random colour. Figure 4 shows a before and after representation of such an image.

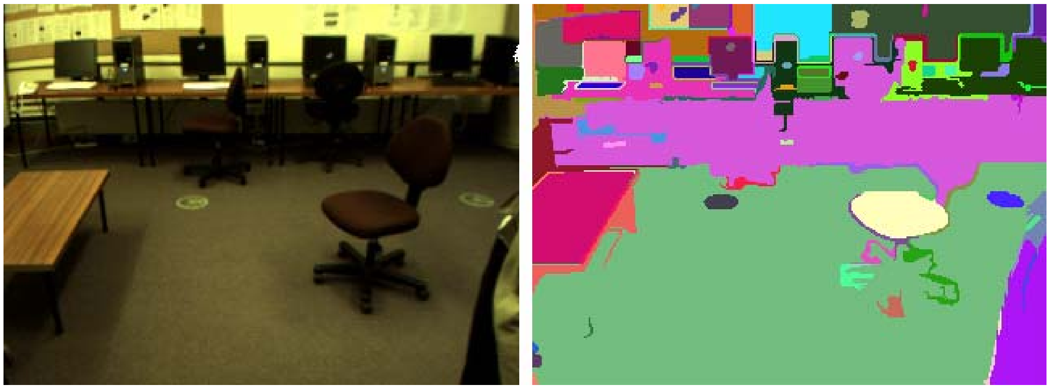


Figure 4. The result of image segmentation on a simple scene.

3.2.2 Region to 3D Position Mapping

The next step is to determine which regions from the segmented image represent which real scene objects/locations. To achieve this, a mapping between the calculated disparity and each pixel in a segmented region is established; pixels for which erroneous or no disparity information could be calculated are ignored. After this process, each region will correspond to a set of pixels and their depth values.

3.2.3 Camera Positioning

In the general case, this new method for plane calculation makes a few assumptions about the position and orientation of the camera to the ground and the scene itself.

- The target search plane is visible to the camera, and bounds the edge of the scene.
- The camera is parallel to the ground plane along both the x-axis and z-axis, therefore perpendicular to the y-axis.

The first assumption is vital because the algorithm will classify the region at the limiting axis direction as a plane regardless. Within this case, the onus is on the user to correctly position the camera so that the plane they wish to locate is visible and bounds the scene (ground plane is the most common).

The second assumption is also critical because if the camera is rotated in any direction the limiting values will no longer correspond to the bounding planar surfaces. The solution is to determine the camera orientation. When trying to determine the position of a bounding planar surface, for example the floor or ground, the average height of each region is examined to discover the limiting value. To ensure that this will always be the case the 3D points that make up the scene must be rectified to a common coordinate frame. This transformation can be achieved by using simple rotation matrices to transform each point back to the reference frame but makes knowledge of camera orientation paramount to the process.

3.2.4 Determination of Plane Parameters

Once the correspondence between the 3D points and regions has been established and the scene has been transformed back to the reference frame, information about the position in the scene of the particular region can be gained by examining the associated 3D data. Generally, any visible bounding planar surface can be found by looking at the region that lies at any particular axis limit. The equation below gives the average position of a region relative to an axis of interest.

$$m_A = \frac{\sum_{i=1}^{N} p_i^A}{N} \tag{4}$$

Where m_A is the mean for the axis of interest for the region, N is the number of points that makes up the region, and p_i^A represents the axis A that is of interest at a point p_i.

For example, if the required plane is the ground plane we will consider only the y-axis and hence y-positions of each region. In the current implementation, the positive y-direction is down. This means the region with the maximum average height (The maximum value of m_A where $A = y$ of each region.) will be considered the planar surface representing the ground. Once the region with the maximum positive y-value is known, we can then deduce the equation of the plane. The general equation of a plane is $ax + by + cz + d = 0$. As the positive y-direction is down, the normal for the ground plane becomes -1 and therefore the equation of the ground plane is $-y + m_y = 0$ where m_y is the mean distance in the y-direction of the region.

3.3 Object Detection

In this project, background subtraction with adaptive backgrounding is used for object detection. This technique is well suited to the situation because of its efficiency and simplicity. Adaptive backgrounding also allows gradual changes in illumination, improving on the robustness of simple background subtraction, lowering reliance on thresholding strategies.

3.3.1 Background Subtraction

Background subtraction is the process of subtracting the current image from that of a reference frame image. That is for every pixel in an image, subtract the value of the corresponding pixel in the reference frame to determine if there is a change at that pixel. Formally, for a single pixel in an image, it can be defined as:

$$D_P = 1 \text{ iff } \left| F_{RP} - F_{CP} \right| > T \tag{5}$$

Where D_P indicates if a pixel is different, F_{RP} is the reference frame pixel, F_{CP} is the current frame pixel, and T is some threshold. The need for a threshold can be attributed to a number of factors being the camera image noise, overall scene changes, and gradual illumination change. Figure 5 illustrates the result of background subtraction

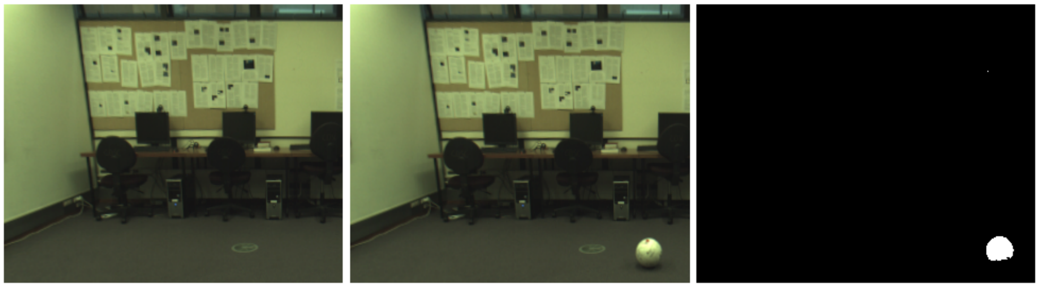


Figure 5. Background Subtraction (The left shows the reference frame, the centre the current frame, and the right the difference image) where only the moving ball is visible in the difference image.

Adaptive backgrounding is a process where by the background image changes at a given time interval or at a given noise threshold. This allows background subtraction to cope with gradual illumination changes and overall scene changes. To illustrate the issue of overall scene changes, consider the difference images of a room if a chair is moved to the other side; every difference frame will now contain both an object where the chair has moved to and another where it has come from. If we replace the reference image currently used for background subtraction with a new image of the rearranged room we will now detect nothing again, which is what is desired. Illumination changes conversely will cause a slow but eventually degradation of the difference image and will require a higher threshold to cope. Updating this reference image periodically allows reduction of the threshold value and therefore improved accuracy for object detection.

Unfortunately, background subtraction suffers from one other main limitation, camera movement. This is the most critical as if the camera is shaking, say outdoors in the wind, much of the scene is likely to be considered moving in each frame.

3.3.2 Object Search

Once regions of interest have been established using background subtraction, it must be determined whether these regions represent candidate objects for tracking or simply noise. A candidate object is defined as an object made up of *N* adjacent pixels from the difference image. A pixel is considered adjacent to another if it occupies one of the surrounding eight pixels. *N* is set to a sensible threshold value that will eliminate noise while extracting candidate objects from the difference image.

3.3.3 Object Elimination

The object search stage may present multiple candidate objects from a single frame. These may be due to noise, shadows or real multiple objects. To be able to track a single object accurately we must attempt to eliminate some of these candidate objects. In this project, elimination of candidate objects is performed using the 3D positional information about those objects. Two of simple heuristics guide this process:

- Objects cannot be further away than any limiting planar surfaces.
- Objects that lie extremely close to planar surface are likely candidates for shadows.

If after these heuristics are applied, no single object is available for tracking, then the conclusion is that multiple moving objects are present.

3.3.4 Object Position

After finding a single object, from a single frame, its position in 3D space needs to be determined. After object search, a set of pixels that represents the object is known. The next step involves determining whether disparity calculations were successful for each individual pixel and if so calculating the average position for the object in 3D space (x, y, z). This is difficult because there are often pixels in the object representation that are part of surrounding objects. Consider the simplified model as shown Figure 6.

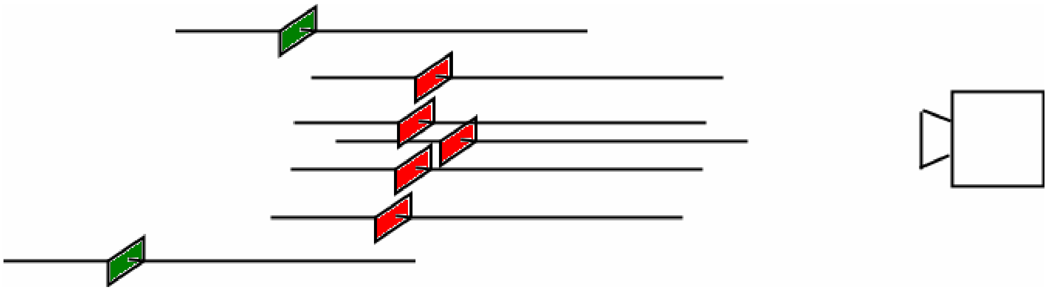


Figure 6: Simplified view of the pixels that make up the object of interest. The red squares indicate pixels that are the object and the green pixels that are part of the background.

As the set of pixels that make up the object should have very similar positions in the disparity map, any pixels with position very distant from the mean should be eliminated. This is achieved by calculating the mean position of an object from the pixel positions, then eliminating pixels that vary by more that three standard deviations from that mean. The mean is then recalculated, and the same process is followed until there is convergence about the true position of the object.

3.4 Object Tracking

3.4.1 Object Trajectories

An object's trajectory is made up of a number of discretely sampled points; 30 per second. This project adopts a model of a single object being visible at any time, reducing complexity. Using this premise it can be assumed that objects in consecutive frames or within some fixed number of frames (threshold) can be considered the same object. The object size (pixel count / distance) across frames proportional to distance from the camera is also thresholded. A brief quantitive analysis showed that this technique proved 94.2% effective in distinguishing between two objects of varying sizes. Objects being registered at incorrect distances caused the largest error.

3.4.2 Calculating Object Trajectories

To determine accurate values for the object trajectory, velocity and position, noise and hence incorrect positional measurements need to be dealt with. Depending on lighting, contrast, and the size of the object being tracked, the accuracy of positional information is significantly affected. Fortunately, as a number of observations of a single object's position are recorded, it is possible to smooth out much of this noise.

Firstly consider an object travelling through 3D space; it will have a continuous position measured with x-y-z values. If we consider the model for the ballistic trajectory of an object, it is clear that while the position is in 3D, its trajectory is constrained to a two dimensional plane, where the plane is defined by its x-z and y coordinates. Figure 7 shows the detected positions of the object, allowing the general plane of travel to be estimated.

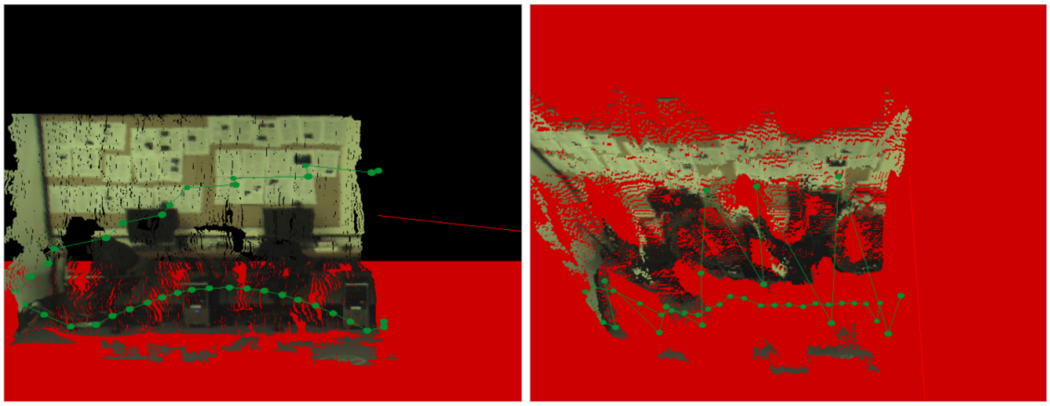


Figure 7: The left image shows an objects trajectory from front on. The right image shows the objects trajectory from top down; the discrete points should appear in the same plane defined by some (x-z) values, but are skewed by noise.

Furthermore, this plane should be deducible from the x-z coordinates of the object's position. The method used to find this plane is linear least squares regression; the result of this methodology is a best-fit line, which together with the y coordinates, defines a plane. The best-fit line is defined formally with the line equation:

$$c_1 z = b_1 x + a_1 \quad (6)$$

The second issue is transforming each recorded object position to the plane. To transform each position to the plane we must first establish the line perpendicular to the best-fit line passing through the point. This can be done by dividing -1 by the gradient (b) of the line, for which the perpendicular line is to be constructed (best-fit line), giving:

$$c_2 z = b_2 x + a_2 \text{ where } b_2 = \frac{-1}{b_1} \tag{7}$$

Subsequently, we need to find the intersection of the two lines. If the determinant of the two lines is less than or equal to zero they do not intersect so we have the following relationship:

$$D = b_1 c_2 - b_2 c_1 > 0 \tag{8}$$

If the determinant is greater than zero than the following equations will give the point of intersection for the two lines:

$$x' = c_2 a_1 - c_1 a_2 \tag{9}$$

$$z' = b_1 a_2 - b_2 a_1 \tag{10}$$

Where x' and z' are the x and z positions of intersection respectively. These points, together with the original y coordinate are used to define the adjusted position of the object.

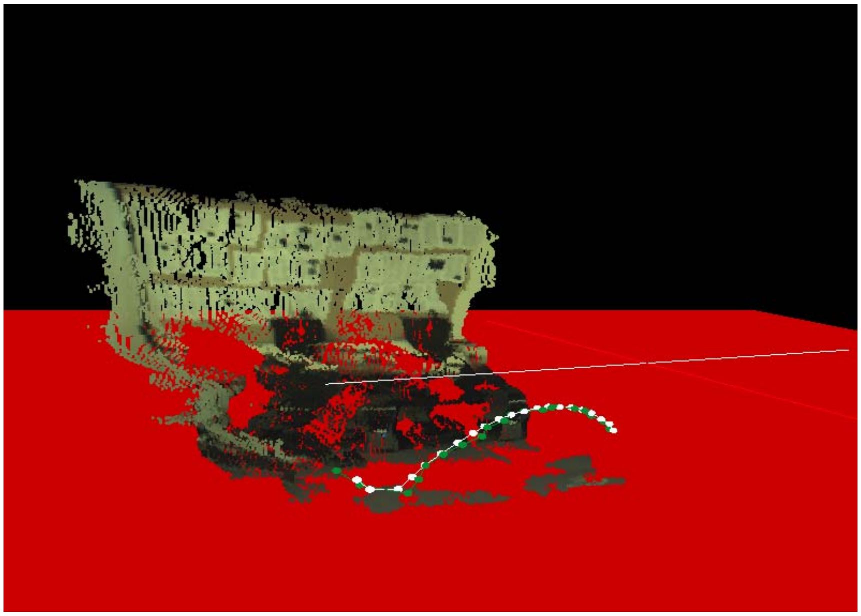


Figure 8: This image shows the points recorded as small spheres with lines joining sequential points. The white line (above objects) indicates the linear least squares fitted line giving direction of travel and the white spheres the adjusted positions of the object.

4 Evaluation and Results

The evaluation in this project focuses on two separate issues, the accurate determination of the ground plane location and the accuracy of object detection and tracking across multiple frames in terms of velocity as a distance measure, and the height above the ground. As the object detection and tracking evaluation is considered relative to the planar extraction, a preliminary evaluation of the planar extraction technique is performed first, followed by a full evaluation of the accuracy of object detection and tracking.

4.1 Apparatus

This evaluation was conducted using a stereo camera system and associated software (Small Vision System) from Videre. It was run on an Intel Core 2 Duo processor with 2GB RAM, and an NVIDIA GeForce 8500GT video card which allowed a frame rate of 30 fps.

4.2 Planar Surface Extraction

4.2.1 Experimental Conditions

To determine the accuracy of this new technique, comparison between physical distance and calculated plane positions is performed for the ground plane location. There are two main influences in calculating the position of a plane, the distance of the camera from the plane and the orientation of the camera relative to the ground. Two different heights examined were 1125 millimetres from the ground and 1445 millimetres from the ground. The first of these heights also had no rotation about any axis, while the second had a 21-degree rotation about the x-axis (camera titled toward the ground).

4.2.2 Experimental Design

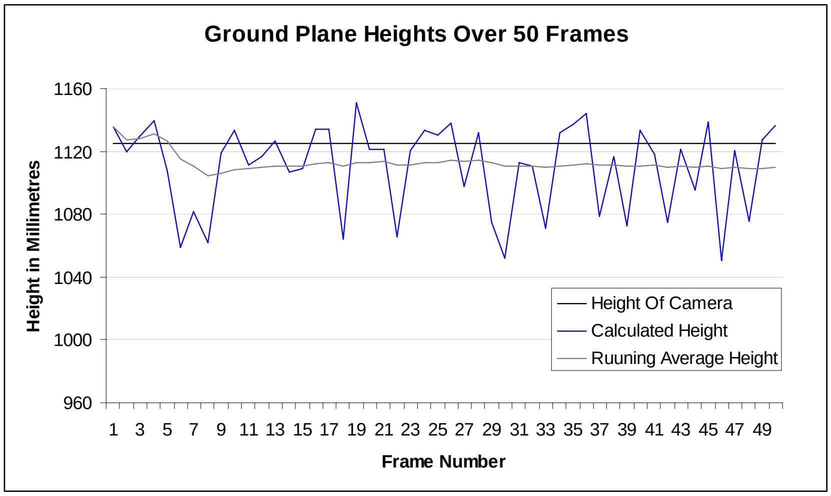
For each height a total of 50 observations were made. This gave 100 observations in total. Observations were made on sequential frames under the same lighting conditions. Each observation resulted in output of an equation representing the ground plane. The point of interest was the distance to the ground from the camera.

4.2.3 Individual Plane Calculation

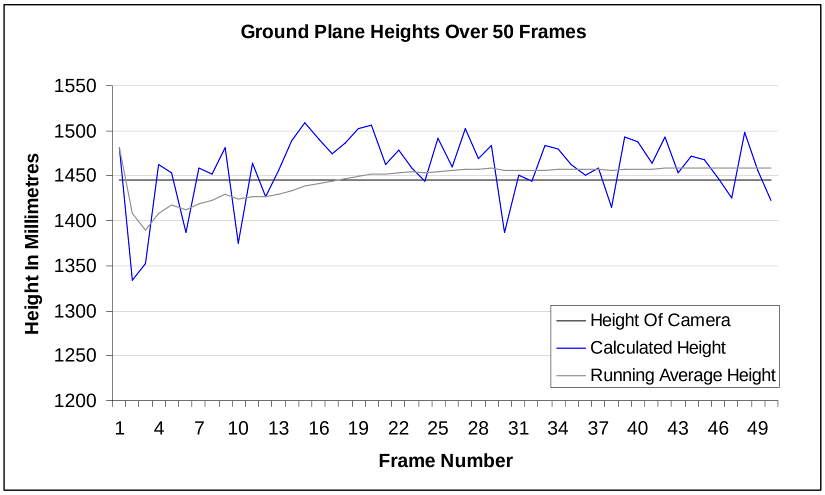
The results for individual plane calculations were accurate for both camera positions with respective orientations. The first camera position with a height of 1125 mm above the ground gave a height range of 101 mm (1050 mm - 1151 mm). The greatest difference between calculated height and actual height was 75 mm from the actual ground plane. The second camera position with a height of 1445 mm above the ground gave a range of 174 mm (1334 mm - 1508 mm). The greatest difference between calculated height and actual height was 111 mm from the actual ground plane.

4.2.4 Planes Across Multiple Frames

The results of the experiment proved to be highly accurate with both heights returning average heights over 50 frames extremely close to the actual height; 1.37% for the camera at a height of 1125 mm with no rotation and -0.90% for the camera at a height of 1445 mm with a 21-degree rotation about the x-axis. Figure 9a-b illustrates these results.



(a.)



(b.)

Figure 9: (a.) and (b.) show the height of the camera (constant), the calculated height of the ground plane in each frame, and the running average height of the ground plane for the two camera positions. Both (a.) and (b.) show convergence of average calculated height to a very close approximation of the actual ground plane.

4.3 Object Recognition and Tracking Accuracy

4.3.1 Experimental Conditions

To measure the object detection and tracking accurately, a number of considerations are included; size of the object, distance from the camera to the object, and the contrast of the object. The most important condition is the size of the object. The size of the object should theoretically allow for easier detection and hence more accurate calculation of position. Subsequently, calculations based on position such as velocity can be used to estimate the overall accuracy of the system. The velocity of the object can be determined by the amount of movement per frame (33ms) and also allows direct comparison with theoretical velocity. The distance from the camera is also important because objects further from the camera will register with fewer pixels and therefore less positional information. Overall the distance will likely affect the accuracy of the positional calculation. Contrast is the last consideration and is important because of the limitations of stereo systems, especially disparity calculation accuracy.

4.3.2 Setup

After detection of an object the position is calculated by the system, but comparing this to real world positional information is made difficult because the position in the real world needs to be reliably measured for realistic comparisons to be drawn. This fact produces a large limitation on the technique for evaluating such a system and means considerable constraints must be put in place to allow for accurate measurements to be taken. In this evaluation, a pendulum type mechanism is used to allow predictable and accurate positions to be found. Figure 10 shows a simple pendulum setup.

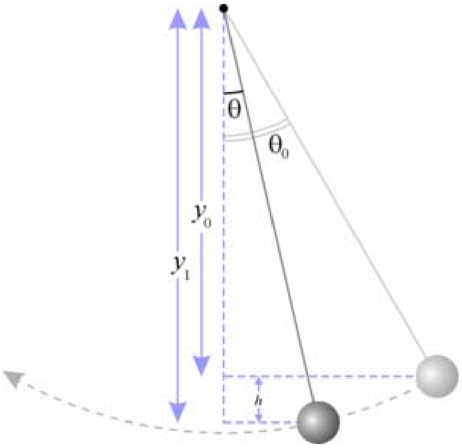


Figure 10: Simple pendulum, y_0 and y_1 give heights at the respective angles θ_0 and θ.

Pendulums are a straightforward physics problem where a mass (commonly referred to as a bob) is attached to a pivot point and is allowed to swing freely. As gravity is the only effect on the mass, neglecting friction, the mass will swing back and forth in a consistent path. By using this constrained system, positions and velocities can calculated at any point in the trajectory of the mass for comparisons against the positions generated by the system. In this evaluation a simple pendulum system [23, 24] is established. The amplitudes or angle from equilibrium is maximised at 15 degrees so that approximation by a simple harmonic is possible. This is feasible as $\sin\theta$ can be approximated by θ, measured in radians, for small values of θ (θ <= 15°), further description is beyond the scope of this project. Formally, the period (time taken for mass to travel left to right and then back again) is defined as:

$$T \approx 2\pi\sqrt{\frac{l}{g}} \tag{11}$$

Where T is the period, l is the distance from the pivot to the centre of mass of the object, and g is gravity (9.8ms^{-2}). Furthermore, the position and velocity of the mass can be defined in terms of the angle and distance from the pivot point to the mass. For this evaluation, two different sized objects are suspended from a pivot with a distance (l) of 2.5 meters to the centre of each. Using the formula in (11) the period should be approximately 3.17 seconds. This period is consistent regardless of the angle, and the weight of the object suspended used for the pendulum bob. The motion of the pendulum bob is governed by the rotational motion principles and is closely correlated with the standard kinematics equations. Rotational motion is defined in terms of θ, the angle measured in radians. This allows deduction of the following rotational motion equations:

$$\Delta\theta = w_i t + \frac{1}{2}\alpha\, t^2 \tag{12}$$

$$w_j = w_i + \alpha\, t \tag{13}$$

Where $\Delta\theta$ gives the change in angle, w_i is the current angular velocity, w_j is the new angular velocity, α is the angular acceleration, and t is the time since the last observation. These two equations allow the calculation of theoretical values of velocity and angle relative to time. Direct comparison between these values and observed values is then possible when determining the accuracy of the system. Table 1 shows the first 10 theoretical values of each to be used for comparison with system calculated values.

θ	w	$\Delta\theta$	α
0.2618	0.0000	0.0000	-1.0262
0.2612	-0.0339	-0.0006	-1.0240
0.2596	-0.0677	-0.0017	-1.0175
0.2568	-0.1012	-0.0028	-1.0065
0.2529	-0.1345	-0.0039	-0.9913
0.2479	-0.1672	-0.0050	-0.9718
0.2419	-0.1992	-0.0060	-0.9481
0.2348	-0.2305	-0.0071	-0.9203
0.2267	-0.2609	-0.0081	-0.8885
0.2176	-0.2902	-0.0091	-0.8529

Table 1: Shown are first 10 theoretical values for angle, angular velocity, angular displacement, and angular acceleration based on a 33ms gap between sampling and gravity as 9.8m^{-2}

Unfortunately, it is very difficult to measure these values using the data from the camera. To circumvent this problem the mathematical relationship between rotational motion and the kinematics equations proves useful. Firstly, the linear velocity of an object is defined as the distance it has moved over a given time period and can be calculated using Pythagoras' rule. Secondly, the relationship between the linear velocity and the angular velocity is given by the equation $v = lw$, where v is the linear velocity, l is the length, and w is the angular velocity from (11). Based on this axiom, direct comparison between observer and actual values can be made. Figure 11 shows the angular and linear velocities of the pendulum bob.

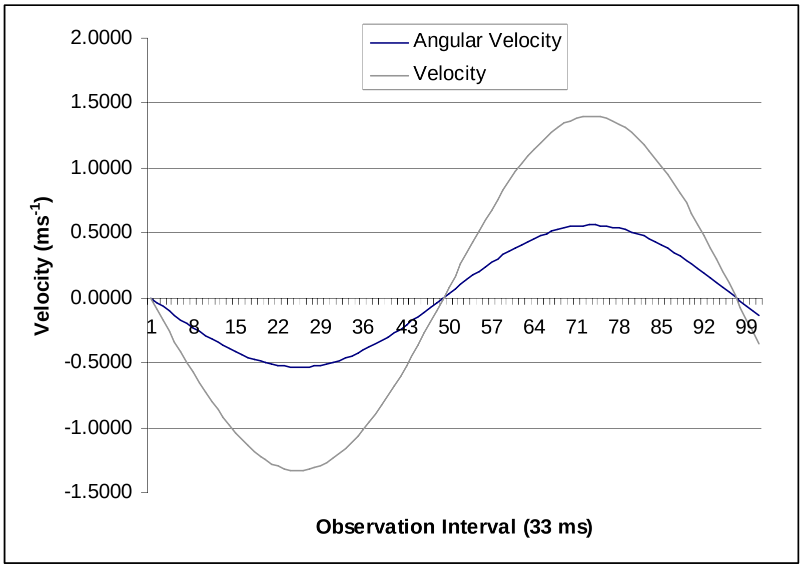


Figure 11: Graph of theoretical angular velocity and linear velocity.

The other value of interest is the position of the centre of mass. The position is defined in relation to the pivot point. Formally its position is:

$$P = l\sin\theta\,\vec{i} - l\cos\theta\,\vec{j} \tag{14}$$

Where P is the position, l is the distance from the pivot point to the centre of mass of the object, θ is the angle from equilibrium, $\vec{i}$ and $\vec{j}$ are unit vectors in the horizontal and vertical directions respectively. The result of this equation gives an x-value representing the horizontal displacement and the y-value representing the vertical displacement. Of these values only the y-value (height) is considered. The y-value is compared with the ground height calculated using the planar location methodology proposed in this paper, to determine the accuracy of both results in relation to real positions.

Additionally, the camera setup is of importance, with the camera being positioned 720 mms from the ground, tilted 15 degrees towards the floor. From the evaluation of the bounding planar surface locating algorithm above, the convergence on the true height appears to have occurred after approximately 20 frames. As the camera was to be in the same position for the entire evaluation the plane's position was calculated in advance, with a calculated height of 722 mms. All positional information is taken with regard to this height.

4.3.3 Experimental Design

The three factors considered in this evaluation are, the size of the object, distance from the camera, and contrast of the object. The two objects had sizes of 90 mms and 220 mms, the distances from the camera were 2 meters and 3 meters, and to illustrate contrast, a white object with high contrast against only parts of the background, and a red object which provided high contrast against the entire background were used. This gives 2 × 2 × 2 = 8 conditions. For each condition, both a single and an average of ten observations were considered. It is expected that a single observation would be difficult to compare with the theoretical positions due to noise in the data. The average, while not entirely useful as an object will normally only be tracked once following a particular path, allows prediction of accuracy based on better object recognition and disparity data accuracy; this is likely the case as systems become more powerful and disparity algorithms become more accurate.

4.3.4 Single Observation

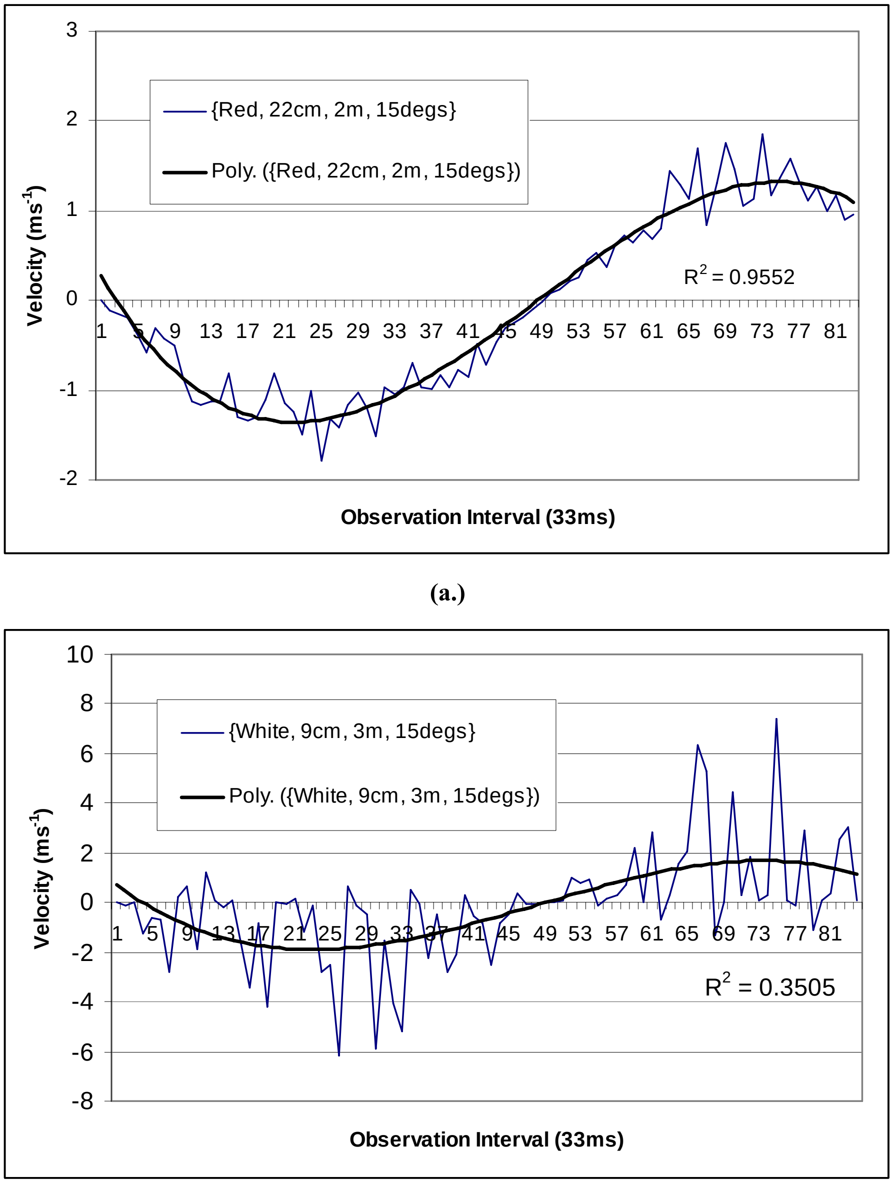
Conditions[1]	Equation	R^2
{W, 9, 3, 15}	$y = -6e^{-5}x^3 + 0.0081x^2 - 0.2824x + 0.0996$	$R^2 = 0.3505$
{R, 9, 3, 15}	$y = -5e^{-5}x^3 + 0.0068x^2 - 0.2274x + 0.7066$	$R^2 = 0.5950$
{W, 9, 2, 15}	$y = -4e^{-5}x^3 + 0.0061x^2 - 0.2043x + 0.5719$	$R^2 = 0.5719$
{R, 9, 2, 15}	$y = -4e^{-5}x^3 + 0.0060x^2 - 0.2065x + 0.7791$	$R^2 = 0.6751$
{W, 22, 3, 15}	$y = -4e^{-5}x^3 + 0.0061x^2 - 0.2142x + 0.6823$	$R^2 = 0.6823$
{R, 22, 3, 15}	$y = -3e^{-5}x^3 + 0.0049x^2 - 0.1693x + 0.4120$	$R^2 = 0.9326$
{W, 22, 2, 15}	$y = -3e^{-5}x^3 + 0.0051x^2 - 0.1739x + 0.4041$	$R^2 = 0.9139$
{R, 22, 2, 15}	$y = -4e^{-5}x^3 + 0.0053x^2 - 0.1823x + 0.4587$	$R^2 = 0.9552$

Table 2: Polynomial approximation equations (order 3) and R^2 value for all conditions from a single observation.

[1] Conditions are shortened and are described as the following: Colour of object (W = White, R = Red), Size of the object (9 = 9 cm, 22 = 22 cm), Distance from the camera (2 = 2 m, 3 = 3 m), Pendulum angle (15 = 15 degrees).

Results from a single observation were varied. Each condition is approximated with a polynomial of order three as appropriate, and visible from Figure 11. Table 2 gives the approximation equation and R^2 value for each of the eight conditions.

Three of the eight conditions, {R, 22, 3, 15}, {W, 22, 2, 15}, {R, 22, 2, 15}, have shown a good approximation with a high correlation value. The other five conditions, {W, 9, 3, 15}, {R, 9, 3, 15}, {W, 9, 2, 15}, {R, 9, 2, 15}, {W, 22, 3, 15}, while having similar equations have shown very low correlation values because of high noise.



(a.)

(b.)

Figure 12: (a.) has conditions red, large object, close distance, with standard angle and represents the best result from a single recording. (b.) has conditions white, small object, far distance, with standard angle and represents the worst result. Note the difference in range on the y-axis between the two graphs.

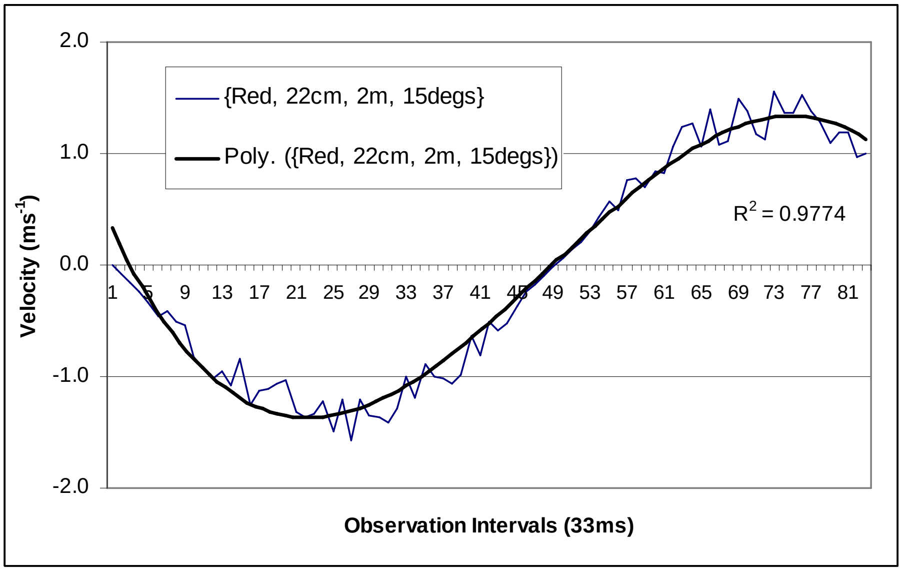
4.3.5 Averaged Observation

Results from the averaged observations were more promising, yet noise is still present across many of the conditions. Each condition is approximated with a polynomial of order three as appropriate, and visible from Figure 11. Table 3 gives the approximation equation and R^2 value for each of the eight conditions.

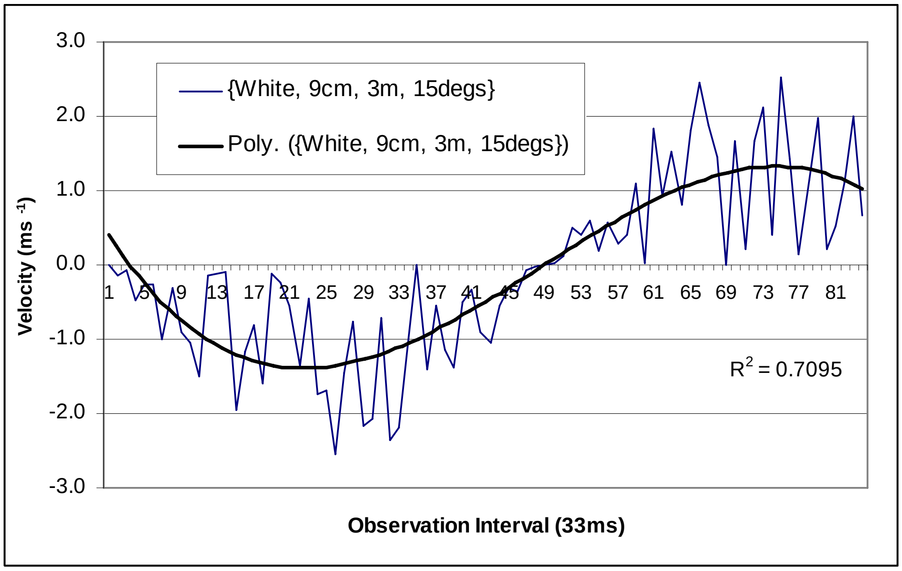
Conditions	Equation	R^2
{W, 9, 3, 15}	$y = -4e^{-5}x^3 + 0.0056x^2 - 0.1953x + 0.5927$	$R^2 = 0.7095$
{R, 9, 3, 15}	$y = -4e^{-5}x^3 + 0.0051x^2 - 0.1723x + 0.4207$	$R^2 = 0.9237$
{W, 9, 2, 15}	$y = -3e^{-5}x^3 + 0.0050x^2 - 0.1715x + 0.4222$	$R^2 = 0.7711$
{R, 9, 2, 15}	$y = -4e^{-5}x^3 + 0.0057x^2 - 0.2003x + 0.6311$	$R^2 = 0.9561$
{W, 22, 3, 15}	$y = -4e^{-5}x^3 + 0.0055x^2 - 0.1892x + 0.5550$	$R^2 = 0.9543$
{R, 22, 3, 15}	$y = -3e^{-5}x^3 + 0.0051x^2 - 0.1745x + 0.4531$	$R^2 = 0.9652$
{W, 22, 2, 15}	$y = -4e^{-5}x^3 + 0.0054x^2 - 0.1857x + 0.4992$	$R^2 = 0.9748$
{R, 22, 2, 15}	$y = -4e^{-5}x^3 + 0.0054x^2 - 0.1868x + 0.5102$	$R^2 = 0.9774$

Table 3: Polynomial approximation equations (order 3) and R2 value for all conditions averaged across ten observation.

Taking an average of ten observations has given a big increase in accuracy with six of the eight conditions, {R, 9, 3, 15}, {R, 9, 2, 15}, {W, 22, 3, 15}, {R, 22, 3, 15}, {W, 22, 2, 15}, {R, 22, 2, 15}, now showing a good approximation with a high correlation value. The other two conditions, {W, 9, 3, 15}, {W, 9, 2, 15}, while having increased markedly, are still showing low correlation values.



(a.)



(b.)

Figure 13: (a.) has conditions red, large object, close distance, with standard angle and represents the best result from an average of ten recordings. (b.) has conditions white, small object, far distance, with standard angle and represents the worst result. Note the difference in range on the y-axis between the two graphs.

4.3.6 Observed Height Results

The calculated height above the ground was also of interest with values roughly proportional to those observed for accuracy. From both a single and averaged observations the heights were similar with heights fairly close to the true height. Figure 14 shows the best result and worst result from height calculations. Other conditions have been excluded for clarity but all fell within the range of the best and worst results.

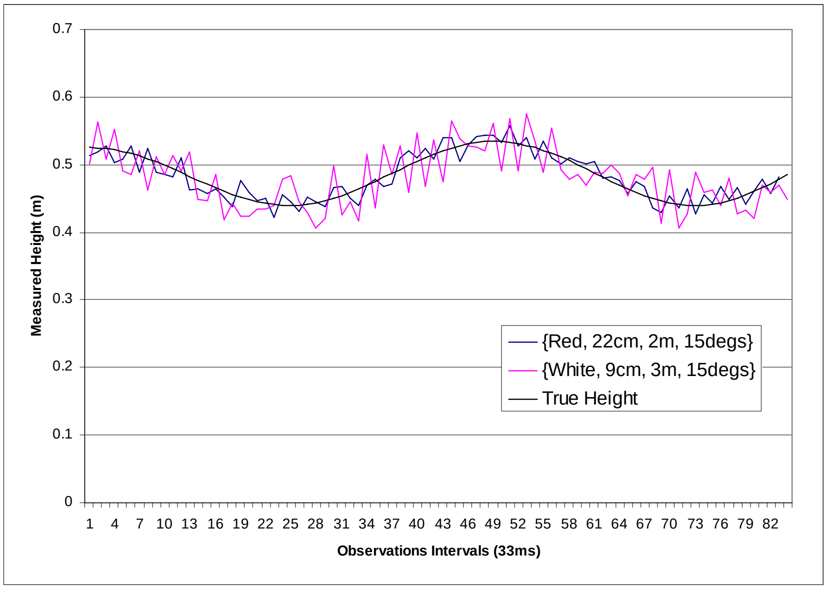


Figure 14: The best condition {R, 22, 2, 15}, and the worst condition {W, 9, 3, 15}, and the theoretical height with respect to time.

5 Discussion

This implementation and evaluation has introduced and shown a new bounding planar surface extraction technique able to locate and find accurate locations of planes. It has also shown that applied statistical analysis of object tracking significantly improves trajectory accuracy.

Bounding planar surface extraction has shown that by using graph-based image segmentation with disparity information, it is possible to determine accurate locations of bounding planar surfaces in a scene, particularly the ground plane. The results show that the calculated plane in each frame will fall close to the actual bounding planar surface and that by using a running average over multiple frames, the planar representation converges on a very accurate location. Accuracies for calculated plane locations have been promising with less than 1.4% difference between them and the actual plane location.

Accurately tracking objects' trajectories has proved difficult, yet this project shows some promising results. Noise has been the main factor causing inaccuracies in both positional and dependent velocity measures. Firstly, single observations were considered but proved inaccurate. To improve accuracy, averages were taken over ten observations. The size of the object, the distance from the camera, and the contrast of the object all have an effect on accuracy. The results clearly show that larger, closer objects with higher contrasts are more effective for tracking. The main reason for this is the amount of information that can be extracted from the disparity data. Further, as all velocities are based upon the calculated position (average position), the number of disparity data points making up that object will always be the principle influence. Distance from the camera will of course reduce the number of disparity points but the following relationships exist.

- The larger the object, the more disparity data points.
- The higher the contrast (which can be aided by higher illumination levels), the more disparity data points.

Furthermore, in the object recognition phase (adaptive background subtraction), a higher contrast will likely result in a larger image difference. This will equate to further disparity data points being used in positional calculations.

There are a number of interesting problems arising from the research and many directions for future work. In the following sections, the process for both bounding planar surface extraction, and detection and tracking of objects are revisited with results and limitations at each stage outlined, together with suggested improvements to planar and trajectory accuracy.

5.1 Planar Surface Extraction

5.1.1 Image Segmentation

In this application, poor image segmentation will have a large detrimental effect on the accuracy of calculation. From Figure 4, it can be observed that an image segment sometimes includes objects with different positions (note the base of the chair). When these situations occur, the 3D positions of unrelated objects will be included in the calculation of plane parameters and reduce the overall accuracy. There is a common trade-off with image segmentation where as we increase the threshold for the degree of difference for neighbouring pixels to accommodate varying textures and lighting (shadows), we will ultimately include pixels that are not part of particular regions.

5.2 Object Recognition and Tracking

5.2.1 Object Recognition

Object recognition is an extensively studied area, resulting in numerous techniques with good recognition results. Adaptive background subtraction was chosen for this project based on its efficient processing time and stationary camera. Object recognition in this project has been sufficient but other techniques could possibly offer better detection. The double difference algorithm, is the most obvious alternative and would increase the robustness of the overall system by increasing resilience to background and lighting changes.

5.2.2 Object Tracking and Trajectories

The object tracking model has one main limitation, which is the assumption that only a single object at a time is visible. This limitation would benefit from positional predictions, to allow multiple objects with differing paths and could also prove useful in reducing the object search space and ensuring the correct object is being tracked. This would have a direct effect on both computation time and overall accuracy given multiple objects. The object tracking information gained from positional information is used to determine trajectories of objects; specifically, the direction of travel and velocity. While the trajectory is calculated using linear regression, and object positions are normalised to this travel direction, results show that noise still has a large effect. Further data fittingof trajectories (such as ballistic trajectories) may provide higher accuracies by further averaging out noise.

Another important observation is that the difference in calculated height to true height is not significantly large. This means the displacement causing the largest positional errors is arising mostly from incorrect depth calculations. The other issue that may cause high inaccuracies is the time interval of 33 ms. Because of this small interval, even slight errors will cause significant errors in positional information.

5.3 General Problems

5.3.1 Stereo Vision

There are two main types of limitations introduced by using stereo vision; scene related issues, such as lighting and occlusions, and technology related issues caused by slow hardware or inefficient algorithms. Scene related issues are the most difficult to solve and usually pose constraints on a system. Solutions to many of these problems are possible with more computationally complex, disparity algorithm implementations. This has been highlighted in both the bounding planar surface extraction and object trajectory calculations by showing better accuracies when considering average positions over multiple observations. Unfortunately, such solutions may well degrade the real-time efficiency of this method. The higher the resolution of the camera, and subsequently the 3D data, the slower the overall system will run and so there is a trade-off between accuracy and running in real-time.

5.3.2 Measurement Accuracy

The position and orientation measurements for the camera and real world objects play a significant role in the accuracy of the system. For bounding planar extraction, measurement of the height of the camera and the orientation (angle in relation to each axis) need to be known to sub-centimetre/degree accuracy in order to provide extremely accurate bounding planar surface locations. The proposed method makes assumptions that air resistance and friction in the pendulum are negligible, and the standard pendulum assumption, that $\sin\theta$ can be approximated by θ (measured in radians) for small angles. Even with these assumptions, the ground truth of exact timings and positional information is manually measured which introduces further errors.

5.4 Future Work

The most immediate research direction is attempting to increase accuracy by using higher image resolutions (640×480) and varying stereo processing algorithms [25]. Overall, this would lead to more disparity points and theoretically better accuracies. The bounding planar extraction methodology could become automated in that the system can find planes with no prior knowledge as to whether they exist or not and no knowledge of camera orientation or position. This may only be possible for complete bounding surfaces.

Future work on object tracking would benefit from multiple object support and predictive position information. Trajectory accuracy would increase if the physics of ballistic trajectories were taken into account. The incorporation of predictive filters such as the Particle Filter would enable search space reduction and multiple object support. Complete object segmentation could be improved using colour to ensure all pixels representing an object have been clustered, resulting in more disparity information. Furthermore, enhanced foreground-background disparity segmentation would be useful in separating objects in the foreground from the background clutter.

6 Conclusion

Tracking a fast moving object with accuracy using stereo cameras would prove useful for many applications; unfortunately it is a difficult problem to solve with current technology. No technique reviewed by this paper or the methodology proposed here can be used alone as a complete and accurate tracking system in all environments.

The first part of the methodology proposed in this project has demonstrated a novel bounding planar surface extraction technique with accurate results within 1.4% of the true plane location. The other part of this proposed methodology, object tracking based on trajectories, has shown mixed results with accuracy dependent on size, contrast and distance of the object from the camera. These issues are all inherent in standard stereo vision disparity algorithms. To get more accurate locations of objects and hence increase the accuracy of a tracking system, the proposed methodology could be enhanced by incorporating predictive filters, colour and foreground/background disparity segmentation, providing strong directions for future research.

References

[1] W. Skarbek and A. Koschan, "Colour Image Segmentation - A Survey," 1994.

[2] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," International Journal of Computer Vision, vol. 59, pp. 167-181, 2004.

[3] R. Jain, R. Kasturi, and B. Schunck, "Machine Vision," pp. Chapter 11, Pages 289-291, 1995.

[4] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," International Journal of Computer Vision, vol. 47, pp. 42, 2002.

[5] R. Michael and M. Jenkin, "Stereopsis Near the Horoptor," Proc. 4th ICARCV, 1996.

[6] B. Liang and N. Pears, "Ground Plane Segmentation for Mobile Robot Visual Navigation," Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, pp. 1513-1518, 2001.

[7] B. Liang and N. Pears, "Ground Plane Segmentation From Mulitple Cues," Second International Conference of Image and Graphics, vol. 4875, pp. 822-829, 2002.

[8] R. Duda and P. Hart, "Use of the Hough transformation to detect lines and curves in pictures," Communications of the AGM, vol. 15, pp. 11-15, 1972.

[9] D. Ballard, "Generalizing the hough transform to detect arbitrary shapes," Readings in computer vision: issues, problems, principles, and paradigms, pp. 714 - 725, 1987.

[10] B. Yu and A. Jain, "Lane Boundary Detection Using a Multiresolution Hough Transform," International Conference on Image Processing (ICIP'97), vol. 2, pp. 748, 1997.

[11] H. Wu, G. Yoshikawa, T. Shioyama, T. Lao, and T. Kawade, "Glasses frame detection with 3D Hough transform," Pattern Recognition, 2002. Proceedings. 16th International Conference on Computer Vision & Image Processing, vol. 2, pp. 346-349, 2002.

[12] S. C. C. Kamath, "Robust Techniques for Background Subtraction in Urban Traffc Video," IS&T/SPIE's Symposium on Electronic Imaging, vol. 5308, pp. 881-892, 2004.

[13] P. Maybeck, Stochastic Models, Estimation, and Control, vol. 1: Academic Press, Inc., 1979.

[14] P. Rosin and T. Ellis, "Image Difference Threshold Strategies and Shadow Detection," Proceedings of the 1995 Britch Conference on Machine Vision, vol. 1, pp. 347-356, 1995.

[15] S. Rougeaux, N. Kita, Y. Kuniyoshi, and S. Sakane, "Tracking a Moving Object with a Stereo Camera Head," In Proc. 11th Annual Conf. of Robotics Society of Japan, 1993.

[16] M. Tanaka, N. Maru, and F. Miyazaki, "3-D Tracking of a Moving Object by an Active Stereo Vision System," Industrial Electronics, Control and Instrumentation, 1994. IECON '94., 20th, vol. 2, pp. 816-820, 1994.

[17] H. Bie, Q. Huang, W. Zhang, B. Song, and K. Li, "Visual Tracking of a Moving Object of a Robot Head with 3 DOF," Robotics, Intelligent Systems and Signal Processing, 2003. Proceedings. 2003 IEEE International Conference, vol. 1, pp. 686 - 691, 2003.

[18] N. Setiawan, S. Hong, and C. Lee, "Multiple People Labeling and Tracking Using Stereo for Human Computer Interaction," in Lecture Notes in Computer Science, vol. 4552/2007. Berlin: Springer Berlin / Heidelberg, 2007, pp. 738-746.

[19] R. Muñoz-Salinas, E. Aguirre, and M. García-Silvente, "People Detection and Tracking Using Stereo Vision and Color," Image and Vision Computing, vol. 25, pp. 995-1007, 2007.

[20] D. Russakoff and M. Herman, "Head Tracking Using Stereo," Applications of Computer Vision, 2000, Fifth IEEE Workshop, pp. 254-260, 2000.

[21] Y. Wakabayashi and M. Aoki, "Traffic Flow Measurement Using Stereo Slit Camera," Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, pp. 198-203, 2005.

[22] V. Kamat and S. Ganesan, "An efficient implementation of the Hough transform for detecting vehicle license plates using DSP'S," First IEEE Real-Time Technology and Applications Symposium (RTAS'95), pp. 58, 1995.

[23] M. Sternheim and J. Kane, "Physics: Third Edition," in Physics: John Wiley & Sons, Inc, 1988, pp. 224-226.

[24] R. Serway and C. Vuille, in Essentials of College Physics, 2007, pp. 337-339.

[25] D. Scharstein and R. Szeliski, "Stereo Vision Research Page," http://cat.middlebury.edu/stereo/data.html, 23/08/2007.