Innovations Project

Richard Clegg - d1169931

March 16, 2007

Abstract

Discussed in this report, is the relationship between 2D and 3D computer graphics during the production of live-action visual effects. An investigation will take place to analyse the similarities and differences between these two categories. I will explore how passing information, specifically depth map imagery, from 3D to 2D graphics effects the pipeline. Emphasis will then be placed on evaluating the uses, importance and methods of capturing depth information in order to achieve believable integration between computer generated imagery and live-action footage.

Contents

Chapter 1

Introduction

1.1 Aims

By researching and writing this report, I aim to achieve the following:

- A better understanding of the relationship between 2D and 3D computer graphics and how information is passed between these two categories. I feel that by understanding a little of the relationship between the two, it will help me to clarify and understand the structure of current working practices and help me to optimize my own project work flow.
- Explore methods and techniques used to help integrate 3D computer generated imagery with live-action footage in a 2D compositing program. This plays an important role in creating believable visual effects and is relevant to my major project. Anything that I can learn will help me tackle the challenges of my major project with more efficiency and professionalism.
- Understand the potential that 2D compositing packages can have and how they can be used to improve productiveness.

1.2 Introduction

This report is split into two main parts. It begins by discussing the relationship between 2D and 3D computer graphics and how information is passed between them. Basic techniques on how to optimise the work flow between 2D and 3D graphics are studied and explained. This starts to reveal 2D/3D cross-overs in situations such as 3D information like Z-depth being accepted as input during the 2D stages of a pipeline. This begins to break down the boundaries between 2D and 3D graphics. The report then concentrates on depth information by looking into depth maps, their uses and their limitations. To provide a 2D compositing package with depth information for live-action footage, the report finally looks at some methods of obtaining Z-depth information from the real world. Finally it is discussed what impact this would have when compositing live-action with CG and if the process of achieving the depth information is worth the results.

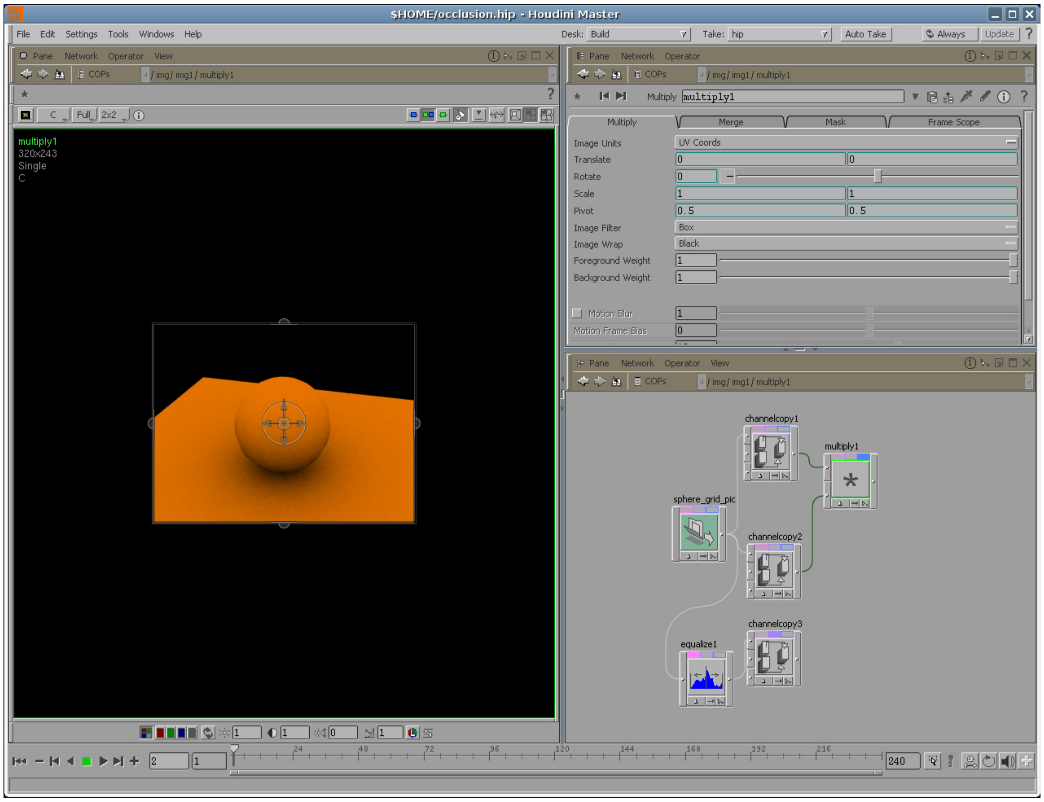
Chapter 2

2D and 3D Graphics

2.1 The Relationship between 2D and 3D Graphics

Computer graphics can be split into two categories: 3D graphics and 2D graphics. 3D graphics is the process of creating a 3D model or simulation of an object or effect in a virtual environment within a computer system. These objects can then be lit, textured and viewed from any angle though an imaginary camera. 3D objects are virtual mathematical representations of real world objects that exist within a computer, and therefore are often treated as such. We can only view these objects through a two-dimensional display device. For the application of visual effects, 3D graphics are usually only the first stage towards the final result. Rendering software is used to output a sequence of images that are then manipulated and utilised in a 2D graphics package to contribute and help produce the final result.

2D graphics traditionally does not attempt to utilise three-dimensional data. Rather it is usually used to enhance and integrate the 3D rendered images and/or real live film or photographs. 2D graphics describes the process of tweaking and layering flat imagery to achieve a final result. Developments in the complexity of 3D and 2D graphics has resulted in gray areas where we are beginning to see 2D/3D hybrid packages[1]. There are now ways of representing 3D data in the form a still image, or as an additional image channel, to allow 2D packages to utilise basic 3D information. This further bridges the gap between 2D and 3D graphics.



Multiply multiply1

Image Units UV Coords
Translate 0 0
Rotate 0
Scale 1 1
Pivot 0.5 0.5
Image Filter Box
Image Wrap Black
Foreground Weight 1
Background Weight 1
Motion Blur 1
Motion Frame Bias 0

Figure 2.1: **3D package Houdini**

Houdini has a 2D compositing section. Here is an example of an occlusion channel being isolated and multiplied on top of the colour channels.

An example of a 3D package that has 2D capabilities can be seen in Figure 2.1. Houdini is broken up into different environments. Figure 2.1 shows Houdini in its 2D compositing environment. In this screen shot you can see an occlusion pass, included as a channel in the image, being separated from the other channels and multiplied on top of the colour channels. This includes occlusion into the colour of the image. Whilst being a useful feature to have, Houdini's compositing capabilities are not as advanced as that of a 2D compositing package such as Shake.

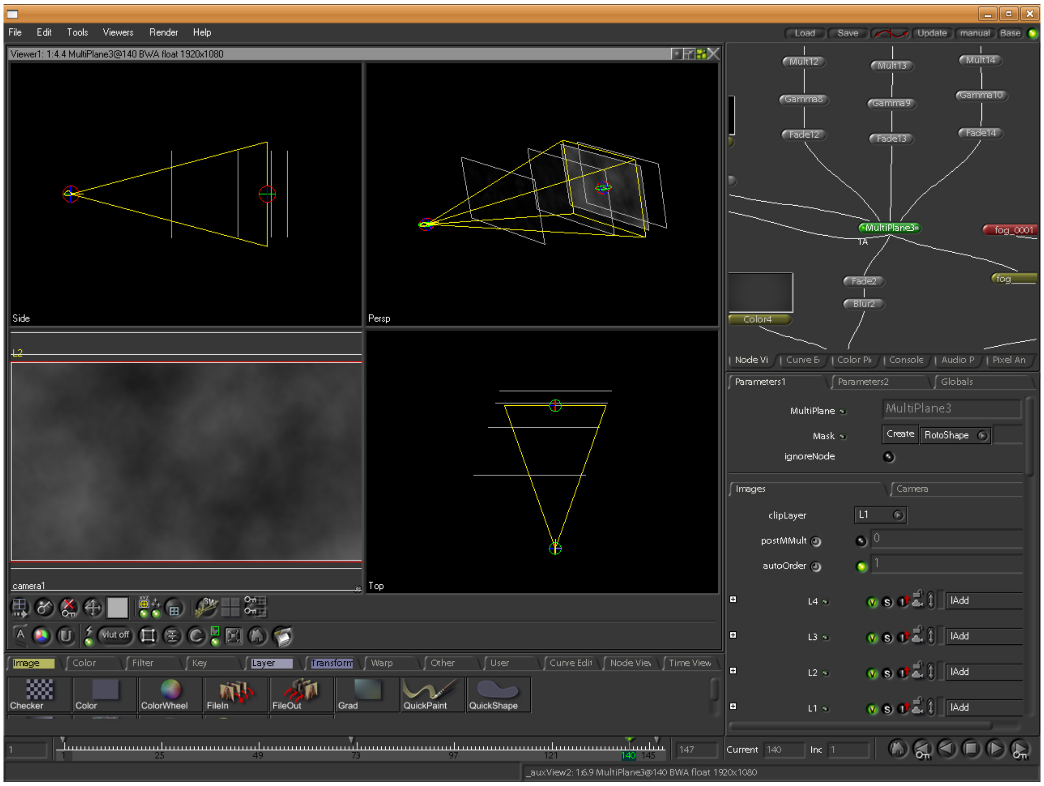


Figure 2.2: **2D package Shake**

Shakes MultiPlane node allows 3D movement of layers as well as an animatable camera.

An example of a 2D compositing package that has 3D capabilities can be seen in Figure 2.2. This is a screen grab of the‘ controls of a node in Shake called the MultiPlane node. This node allows the 3D movement of layers in 3D space. This can then be viewed through a 3D camera, which itself can be moved and animated. Again, this is not as sophisticated as a 3D package alternative, but it is an extremely useful tool that enhances the power and control that Shake can have. This is a good example of a 2D/3D hybrid program.

2.2 Rendering In Layers and Passes

Although it is possible to render a 3D scene as a final finished shot. In order to composite the elements of a shot together more effectively, it is common practice when rendering 3D graphics, to break the scene down into separate layers for each element. This is called rendering in 'passes' or 'layers'. Rendered passes and layers can then be brought together in a 2D compositing package to compose the final shot:

> "This methodology allows the 3D artist to avoid time-consuming rerenders of large scenes when there is only a problem with a specific element. Instead, only the element in question will need to be redone, usually saving a great deal of time and resources."[1]

2.2.1 Rendering In Layers

Rendering in layers is the process of rendering different objects or groups in a 3D scene separately[3]. An obvious example would be to separate the foreground objects from the background.

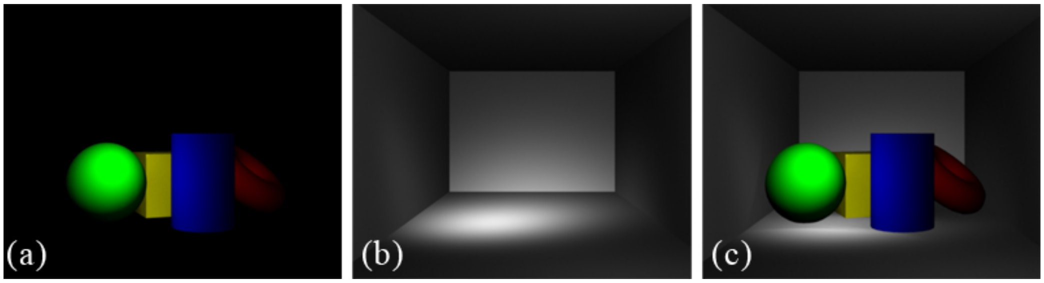


Figure 2.3: **Render Layer Example.**

a) Object layer.
b) Background layer.
c) Layers a and b composited together.

Figure 2.3 demonstrates this. Here you can see that Figure 2.3a is a render of some basic primitive shapes. Figure 2.3b is a render of a background room for the objects to sit in. Figure 2.3a is rendered with an alpha channel, so that when layered on top of Figure 2.3b, you have a finished scene (Figure 2.3c). However, there is now an additional level of control that allows each layer to be adjusted in a 2D package, rather than performing any potential time consuming re renders to achieve the desired result.

2.2.2 Rendering In Passes

Rendering in passes is the process of rendering different attributes of your scene separately. This is especially useful for adjusting the lighting in the 2D stage.

> “By combining and adjusting different passes in a compositing program, a scene can be tweaked interactively without being re rendered, and subtle effects can be precisely fine-tuned or matched to a filmed background plate.”[1]

Rendering out in passes allows a more interactive approach to fine-tuning and scene adjustment in a 2D package. This is far more efficient than attempting a trial and error approach such as tweaking and re rendering in a 3D application.

A 3D rendering program will calculate different contributions of a scene that effect the colour and shading. It then layers up all contributions to produce the final render or ‘beauty’ pass (Figure 2.4e). It is possible specify for these contributions to be output on their own, as passes, before they are culminated into one beauty pass. Figure 2.4 lists the main passes that contribute to a beauty pass.

a) Shadow pass - This render provides shadow information where objects block one another from a light source. This can then be used to darken or reduce the colour information received in areas that are supposed to be in shadow.

b) Reflection Pass - Some objects can have shaders assigned to them that will receive reflections that are calculated via ray-tracing. All colour information that is received from the reflected rays are stored in this pass.

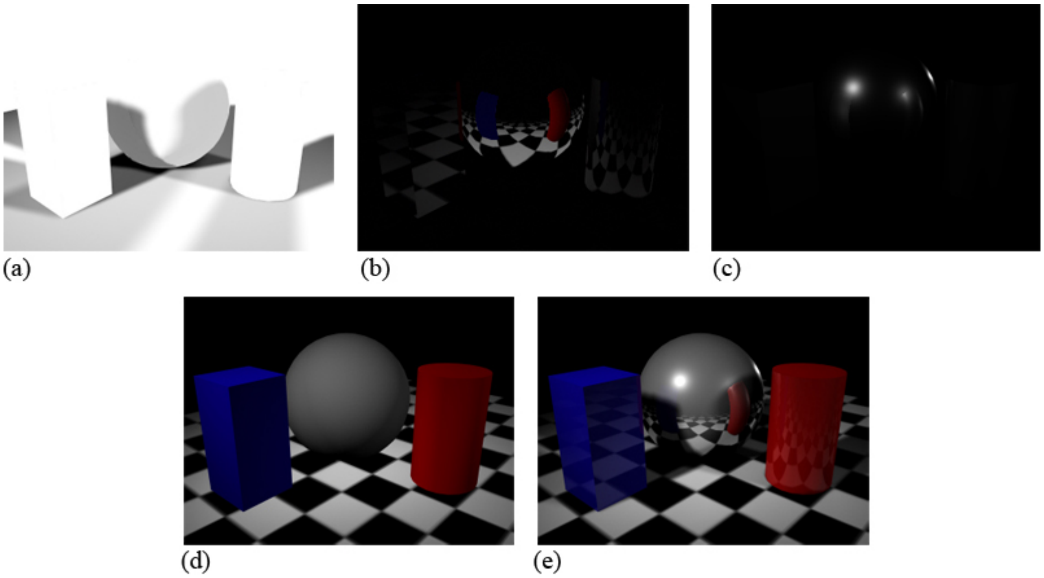


Figure 2.4: **Common pass examples.**
a) Shadow pass.
b) Reflection pass.
c) Specular pass.
d) Diffuse pass.
e) Beauty pass (The culmination of the previous four passes. This is a render of all the elements together).

c) Specular Pass - This pass contains all colour contributions received from the specular highlights off shiny objects in the scene.

d) Diffuse Pass - This pass holds all the colour information received from only the diffuse lighting. Specular highlights, shadows or reflections are not included in this.

Figure 2.5 shows examples of passes that are not always required, but are commonly used for specialised reasons. Figure 2.5a shows an ambient occlusion pass. Ambient occlusion is a cheat that simulates what we sometimes see in the real world. Usually 3D rendering does not simulate real world ambient light that is created when light bounces off one object onto another. This creates the effect where an object's surface that is in cracks, corners or close to other object surfaces, appears to be in darker shadow than the rest of the scene. An ambient

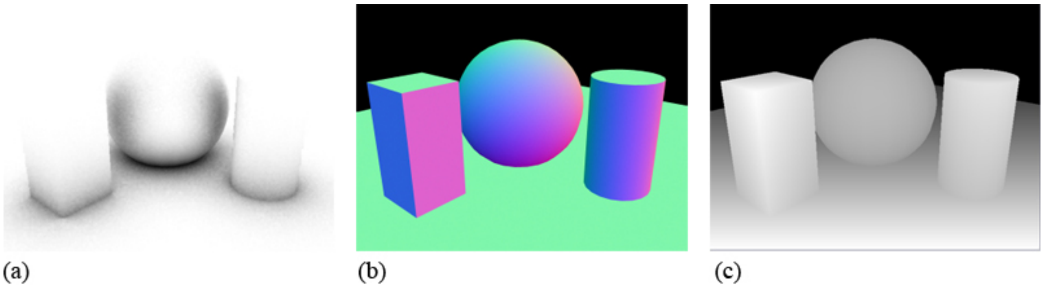


Figure 2.5: **Additional pass examples.**
a) Ambient Occlusion pass.
b) Normal pass.
c) Depth Map pass.

occlusion pass is a form of shadow pass to produce this effect and can be multiplied on top of the beauty pass to shade the necessary areas of the scene.

Every object in a 3D scene has a colour, material and illumination assigned to it. To calculate the lighting, every face on the object must have a normal which describes the direction that the polygon is facing. Normal passes such as Figure 2.5b, are rendered image files that contain information about the direction of the normals on the faces of the 3D objects. Each colour channel in the image represents a directional axis. For example, the red, green and blue channels represent the X, Y and Z axis respectively.

When an image is rendered, the 3D software also calculates the spacial relationship between objects in the scene. This is so that it can decide how objects occlude each other. This information can also be graphically represented in the form of a depth map (See Figure 2.5c). Depth maps only require a single channel where each pixel value represents its distance from the camera. Generally, darker areas suggest that the the pixel is further away from the camera, where as lighter pixels suggest that the pixel is closer. However, depending on the software used, it is sometimes the inverse of this. Depth maps are sometimes referred to as Z-depth images. Some possible uses of depth maps and normal passes will be discussed later in this report.

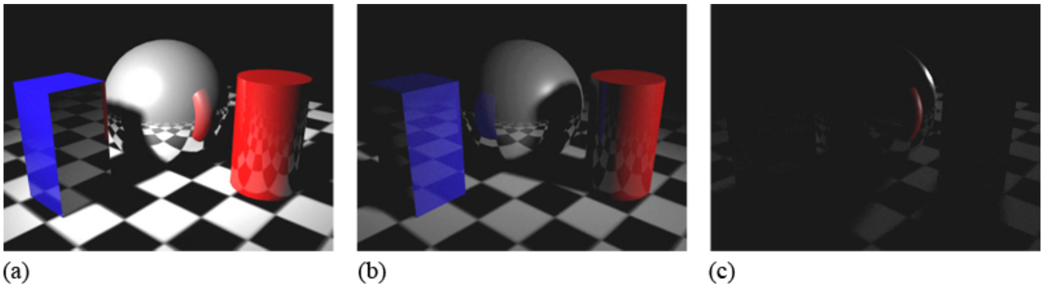


Figure 2.6: **Light Pass Examples.**
a) Key light pass.
b) Fill light pass.
c) Rim light pass.

It is common practice to render different lights or light groups out in different passes. For demonstration purposes I used a 3 point light setup in a simple scene with primitive geometry. Figure 2.6 shows all three lights rendered out as separate passes. These can then be added together in a 2D compositing package to create a beauty pass like that in Figure 2.4e. Rendering the lights out in passes adds a great deal of flexibility and control to the compositing process[3] as the individual lights can be tweaked and adjusted interactively in a 2D package, without performing time consuming re renders. It also helps the artist to gain a better understanding of how the scene is illuminated, as it is possible to see the influence that each individual light has on the scene.

Rendering in layers and passes can also be beneficial for the following reasons:

- More freedom and control is given to the 2D compositors to make adjustments to the individual elements of the shot. Some of these adjustments such as colour correction/matching, brightness and contrast, would normally be far more time consuming to achieve in the original 3D renders than on individual passes in 2D.
- Rendering in layers also proves useful when the scene has a static background. Only one image of of the background requires rendering, rather than an entire sequence. A rendered image sequence would only be necessary for any moving elements in the shot.

- Different layers may also require different rendering quality settings. By rendering in layers, you can vary quality settings for features such as motion blur, anti-aliasing or ray-tracing, which can all be time consuming if high, on a per layer basis. Where as if the scene were to be rendered as a whole, then every element in the scene must use the same high settings that actually only one portion may require.
- To achieve a depth of field effect, it is far quicker to blur a background pass in 2D than it is to use an expensive depth of field effect in 3D.
- Elements that are rendered in passes or layers can be recycled or reused. For example, a render layer of an effect such as smoke and dust can be overlapped and modified to create a more dense and complex effect.

> "Almost every rendering system has limitations that are best resolved by rendering in layers ."[3]

Rendering in layers and passes generally increases flexibility and levels of control that 2D compositing packages can have on a scene. This in turn provides alternative 2D solutions to problems that would originally be tackled in 3D. This raises the question "Do we utilise the 2D solution or the 3D solution?".

2.3 The 2D or 3D Solution?

During the compositing process, when working with passes, if an element or pass needs tweaking, there are two options. The first would be to make adjustments in the 3D side and re render an updated pass. The second option is to modify the existing rendered pass in 2D. The decision on which option to choose can depend on how complicated the problem is, and the amount of time available.

> "Rarely is the decision a trivial one, and often you will find that the best method is to attempt the changes first in 2D and then revert to the 3D solution as needed."[1]

Solving a problem is usually initially tackled in 2D before 3D because of the speed of the process. There is far more data and information to processed in a 3D package, making it a slower solution. 2D software at its most basic, is simply dealing with arrays of pixels. As 2D software scenes become more complex and intricate, there is usually a noticeable change in the time it takes to process tasks. However it is still much faster, and therefore far more interactive than 3D software. This is why it is a preferred choice. However, it is not always possible to achieve the same results in 2D, and so it is necessary to go back to fixing the problem in 3D.

For example, suppose the lighting of a shot required altering to create a specific mood. With the lights rendered out in passes, It may be possible to achieve the desired effect by simply tweaking the colour values, brightness levels or masking out portions of the existing light in 2D. However, if a more dramatic change needs to be applied, such as the direction of the light source, it would be far easier to implement this in 3D than it would in 2D. The artist can easily move and transform a light in 3D space and therefore more drastic changes will usually require the 3D solution and a re render.

> "Some of the more sophisticated software packages on the market will actually offer the user a hybrid 2D/3D solution, wherein certain parameters that would normally only be available in a 3D package are stored in a specialised image file."[1]

A normal pass and a depth map image are both obvious examples of a 'specialised image file'. Normal passes contain 3D information about the direction the polygon surface normals are facing. Depth maps contain 3D information about the distance a pixel is from the camera. Being able to output three dimensional information into a two dimensional image, therefore making the data a valid input for a 2D package, gives new levels of control to the compositing process. Now that the 2D package is handling three dimensional data, can it still be classified as 2D? As discussed earlier, this is an example of a 2D/3D hybrid situation sometimes referred to as 2.5D.

Depth maps play an important role in expanding the capabilities of a 2D package.

Chapter 3

Depth Maps

> "By incorporating depth, powerful additional editing is possible, as well as changing the camera viewpoint"[8].

3.1 Examples of 2D Uses For Depth Maps

3.1.1 3D Relighting in 2D

With a combination of a normal pass and depth map, it is possible to make basic 3D light adjustments interactively in a 2D package. Whilst researching for this project I discovered a macro for shake which is available to download off the internet [5]. The macro is called NormalLight3D. If supplied with a normal pass and a depth map image, it is possible to perform real time 3D lighting within the 2D compositing package Shake (Figure 3.3). The images in Figures 3.1, 3.2 and 3.3 are screen shots taken from an example shake scene file that I downloaded with the macro. This macro was developed with standard shake nodes, and with enough skill and knowledge could be extended and made even more versatile.

> "The NormalLight3D node takes in a normal map and a z-depth channel and will return 3D lighting information. It uses an accurate vector dot product lighting method and

will return diffuse lighting and Specular highlights. Features include falloff, specular highlights, color channel input for compositing, colored lighting and a 3D light with quality settings."[5]

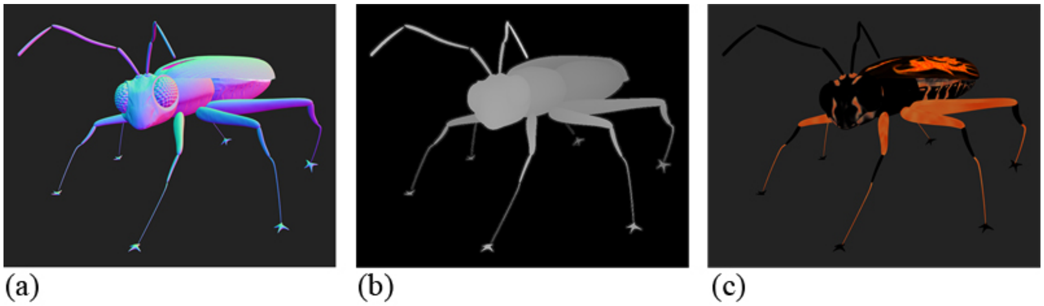


Figure 3.1: **Inputs required for 3D lighting in shake[5].**
a) Normal Pass[5].
b) Depth Map[5].
c) Colour Pass[5].

Using the depth map and colour map with an optional colour pass (Figure 3.1), you can position and manipulate 3D lighting interactively with a 2D compositing package (Figure 3.2).

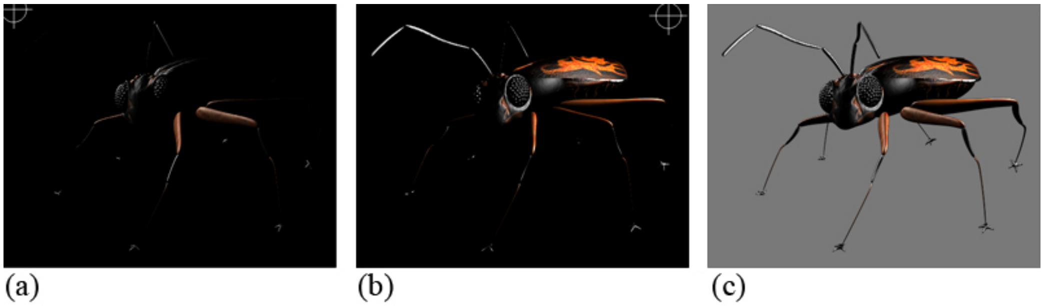
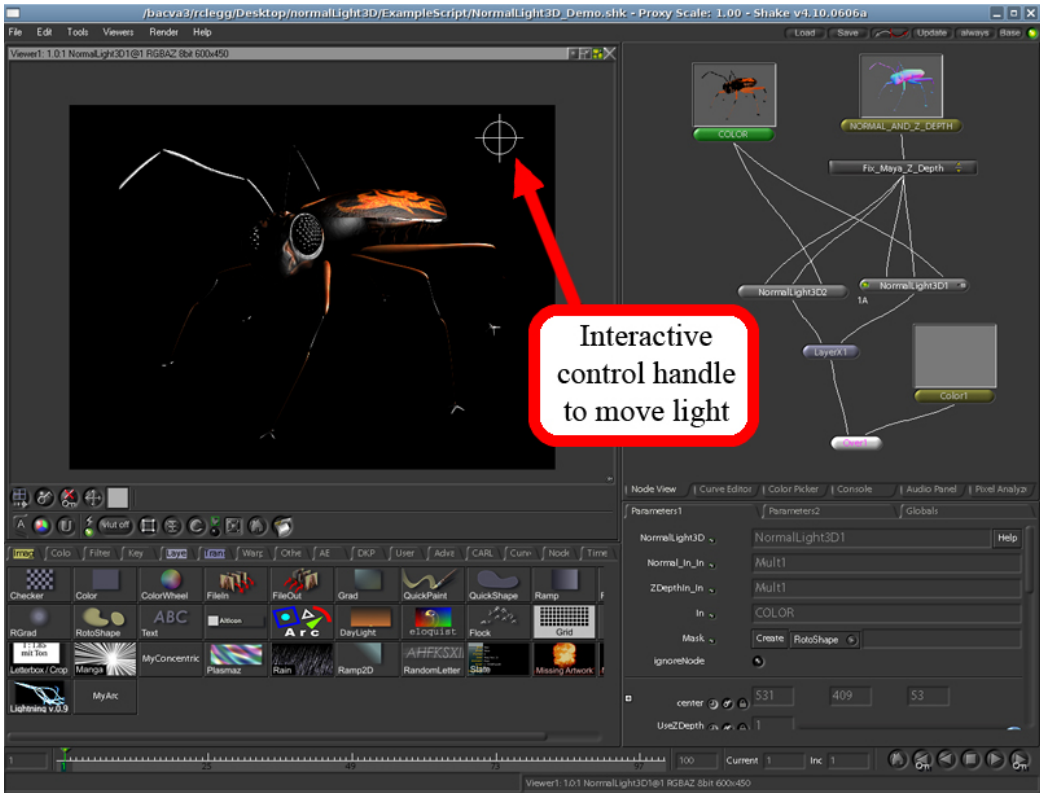


Figure 3.2: **3D lighting performed in 2D compositing package Shake[5].**
a) Light one colour contributions, with light control visible in the top left corner[5].
b) Light two colour contributions, with light control visible in the top right corner[5].
c) Final image lit with two 3D lights. Overlaid onto gray background[5].

This would be incredibly useful to make slight additions or adjustments to the lighting in any live-action footage. However gathering useful



Interactive control handle to move light

NormalLight3D: NormalLight3D1
Normal_In_In: Mult1
ZDepthIn_In: Mult1
In: COLOR
Mask: Create RotoShape
ignoreNode
center: 531 409 53
UseZDepth: 1
Current: 1 Inc: 1

Figure 3.3: **A screen shot of the compositing package shake performing 3D lighting with the NormalLight3D macro[5].**

depth information for live action footage is difficult, as will be discussed later.

3.1.2 Fog and Atmospheric Effects

Often the mood and feel of an environment will be altered during the compositing stage. Depth maps can be useful for adding depth-based atmospheric effects[4]. With the use of a depth map, colour corrections that intensify or vary with distance can be made. This can be used to help enhance the sense of depth that an image can have. Atmospheric fog is a good example of such an effect.

3.1.3 Depth of Field and Defocus

Real world cameras have a depth of field which is the area in front of and behind the focal point that appears to be in focus through the camera lens. Any objects outside of the depth of field appear to be blurred in the image. Depth of field is used creatively to put the object of interest in sharp focus and the background out of focus in order to focus the audience's attention on the subject[6]. In 3D graphics, this can be simulated at render time as part of the output images. This is often time consuming and cannot be easily altered at a later stage. However by using a depth map, defocus effects can be interactively altered and adjusted far quicker and more effectively in a 2D package.

3.1.4 Object Occlusion and Z Depth Compositing

Depth maps provide accurate information about the depth of the objects in the scene. This means that it is possible to determine which objects occlude others with great ease. In a less complicated scene (Figure3.4), this may not be a large advantage, as masks for objects can be quickly keyed to achieve the same result. The benefits of using a depth map start to become more obvious as a scene grows in complexity and becomes dense with objects.

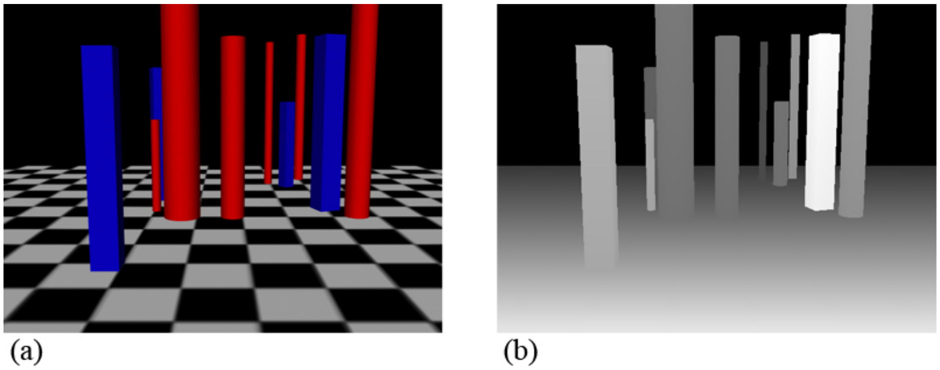


Figure 3.4: **A simple 3D scene.**

a) A beauty render of the scene.

b) The Depth Map image for this scene.

If we have another scene such as Figure 3.5a, which also has a depth map (Figure 3.5b), we can combine the two images (Figure 3.4a and Figure 3.5a) together using Z-depth compositing. Using the depth maps for each of the images, the software can automatically determine the proper foreground/background relationships for each of the different objects in the scene[1]. Figure 3.5c shows the result of this process.

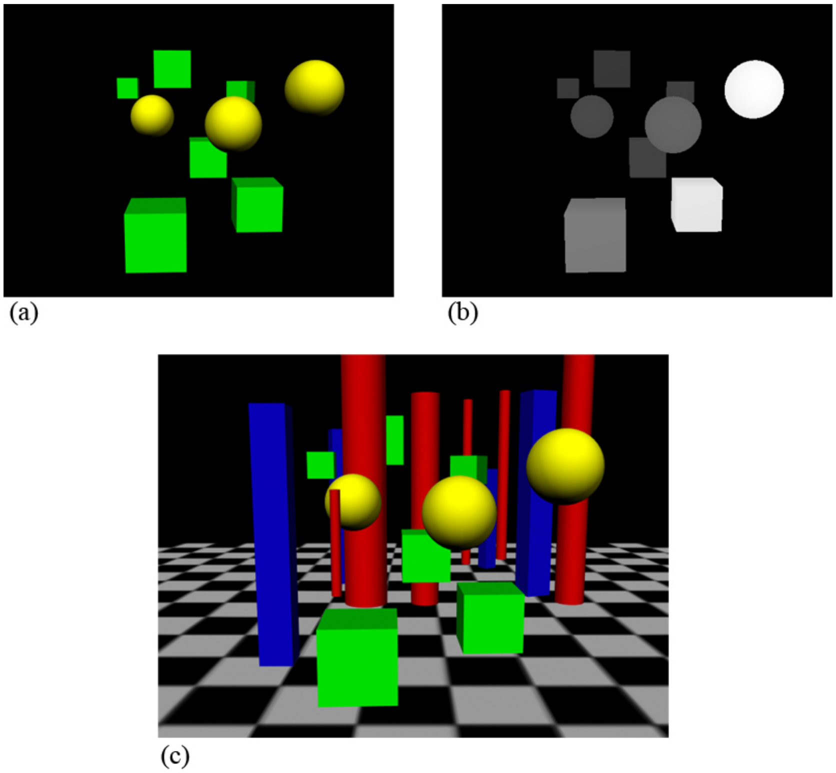


Figure 3.5: **Z-depth compositing.**

a) A beauty render of a second scene.
b) The Depth Map image for this scene.
c) The Z-composite of Figures 3.4a and 3.5a.

The same process done manually would have take a great deal of time to individually determine which objects should overlap.

3.2 Limitations of Depth Maps

3.2.1 Transparency

Each pixel on a depth map image represents a distance from the camera. Z-depth compositing uses this to work out which objects overlay each other and mask each other out. If we have transparent or semi transparent objects in the scene, the pixels for our colour image will receive colour information from both the foreground object as well as from any object that is behind this object[1]. It is not possible to assign a depth value that will represent this.

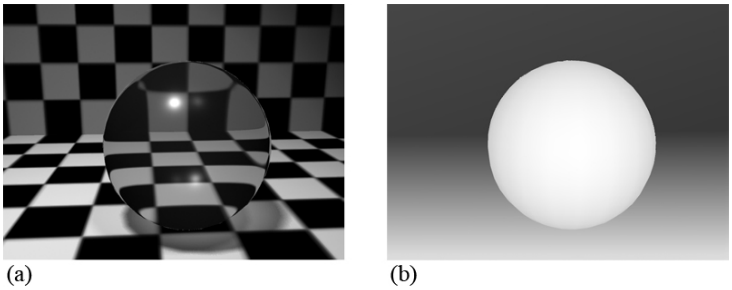


Figure 3.6: **A simple scene with a semi-transparent object.**
a) The colour render of the scene.
b) The depth map image for this scene.

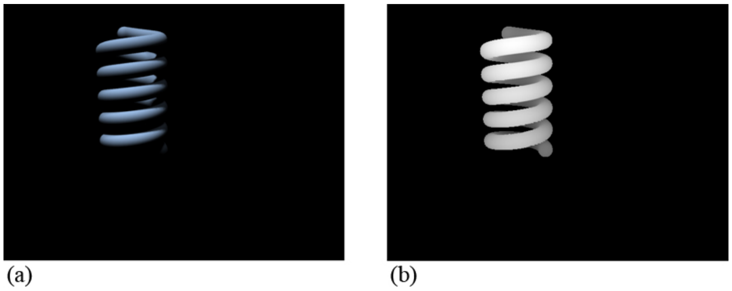


Figure 3.7: **A helix shape to be Z-composited with Figure 3.6a**
a) A render of the new shape.
b) The new shape's depth map image.

The following example and diagrams are based upon examples found on page 258 of 'The Art and Science of Digital Compositing'[1]. In Figure 3.6a we can see a simple scene with a semi-transparent sphere over a checkered background. Figure 3.6b is the depth map for this image. If we use the depth map of this foreground object to Z-composite Figure 3.7a over Figure 3.6a, then we see the results shown in Figure 3.7a. This is wrong as no colour information from the helix, which would be visible behind a transparent sphere, can be seen when the sphere is in front. Figure 3.7b shows what the final image would look like if the two elements were integrated properly. Unfortunately, there is no way to fix this problem, short of re rendering the scene as an integrated whole (without layers) or rendering all the elements individually[1].

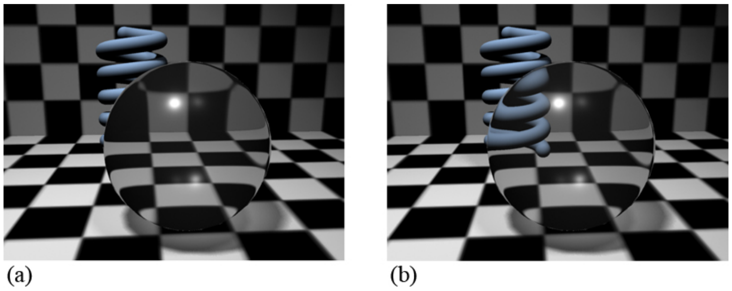


Figure 3.8: **Compositing Figures 3.6 and 3.7**

a) An unsuccessful compositing attempt based on depth information.

b) Proper integration of the two elements by rendering the scene as an integrated whole.

3.2.2 Anti-Aliasing

True depth maps that are output by a renderer do not feature anti-aliasing (Figure 3.9b). This is because each pixel can only represent one distance. This causes a problem as most renderers render out objects with anti-aliased edges. Anti-aliased edges effectively adds a small amount of transparency around the edge of the object (Figure 3.9a). A pixel that is on the boarder of an object will receive colour information from both the object itself and part of the background.

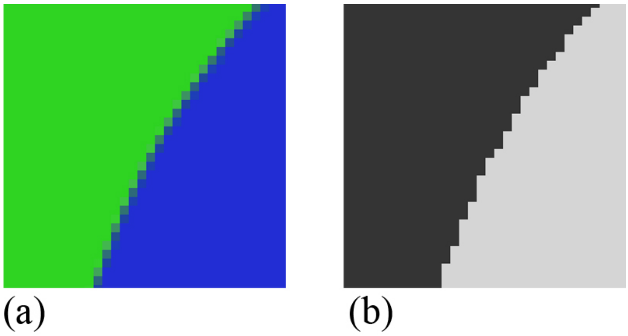

(a) (b)

Figure 3.9: **Close up of a colour render and it's depth map.**
a) Green background against blue object with anti-aliasing.
b) Depth map for the image. No anti-aliasing.

3.2.3 Precision

There is no standard that specifies what exact distance corresponds to what exact pixel value. This is impossible as two different scenes may have massively different scales. Normally this is governed by the user specified clipping plane or clipping boundary that the camera has. A clipping plane or clipping boundary is the distance from that camera that objects are visible to the camera. Any objects beyond that distance will not be visible in any renders. The pixel values are then normalised to the clipping boundary distance. For one scene the clipping boundaries may represent only a few metres, where as in another they may represent several kilometres. To resolve this, it is possible to normalise to depth maps to represent the same distance, however this takes up more time and resources and so reduces the benefits of depth maps in such situations.

The bit-depth of an image also effects the precision. an 8 bit image only has 256 different colour depths to represent the distance from the camera to the clipping boundary. Floating point images obviously increase the precision as well as the flexibility for the range of depth values that are used[1].

3.2.4 Live Action Footage

Z-depth compositing is primarily used with computer generated images, not live-action footage, as it is extremely difficult and very uncommon to be able to automatically generate Z depth information for that medium.

> "It is important to recognise that there are several limitations to Z-depth compositing. The most obvious is that this Z-depth information is generally unavailable for any live-action elements."[1]

Visual Effects usually involves integrating live-action footage with computer generated imagery. For the many advantages that depth maps would provide to do this, depth information is sometimes faked or made up to match, and be used in conjunction with live-action footage. Sometimes it is hand painted, but this requires a great deal of skill and time. An alternative is to use 'dummy geometry' in a 3D virtual environment, that represents the objects live action plate, from which a depth map can then be generated.

> "Research is being done on allowing Z-depth information to be automatically extracted from live-action footage. Stereoscopic photography inherently contains some depth cues, and even single camera view can give some information if it is moving enough to produce parallax shifts. Sophisticated software algorithms are able to analyze such imagery and can often produce some useful depth information for the scene. As the availability and reliability of these tools increases, expect to see more and more use of Z-depth compositing in with live-action footage."[1]

Chapter 4

Depth Information For Live-Action Footage

Having a depth map can be extremely useful for 2D compositing and visual effects. Whilst very simple to obtain for computer generated imagery, it is difficult to acquire for live-action footage. Film and Television visual effects often combine both computer generated images with live-action footage. It would therefore make sense to have the same information for both elements for a more precise integration of the two. Part of the research for this project involved looking at possible ways that depth map images could be extracted from a live-action environment or created from live-action footage. There is currently no standard way or general technique that is used to achieve this. There are many different methods and varying options, all of which have their advantages and disadvantages. The development of numerous ideas and technologies to extract depth information, emphasises the potential usefulness of such data.

With so many varying ideas, algorithms and technologies, it is impossible to describe them all. Many of them are simply varying approaches to similar techniques. Through investigation, there appears to be two main categories that all the different methods. The first category can be referred to as 'Light Detection and Ranging'. This attempts to extract depth data from the live-action environment, utilising some form of special equipment. The second category can be referred to as 'Image Based Modeling'. This attempts to create depth information from depth cues seen in recorded footage or video.

4.1 Image-Based Modeling

4.1.1 What is Image-Based Modeling?

> "Recently, creating models directly from photographs has received increased interest in computer graphics. Since real images are used as input, such an image-based system has an advantage in producing photorealistic renderings as output."[9]

Image-based modeling is the process of using complex mathematical techniques to calculate and build virtual 3D models of an environment from real photographs. This approach to obtaining depth information, involves reconstructing the live-action environment virtually, sometimes called building dummy geometry. Using the representational 3D models of the environment we can then obtain depth maps in exactly the same way we would for any other 3D model. There are many different algorithms and attempts to achieve this. Some systems only require a single photograph to calculate a 3D model. However the most successful and more commonly used approaches use stereoscopic imaging. This is when more than one photographs are taken from different angles of the subject in order to calculate depth, working in a similar way to the human eye. This is the approach that I will be discussing.

4.1.2 How Does it Work?

Stereopsis

Stereopsis is the process of perceiving depth from stereoscopic imaging or stereoscopy[11]. It works in much the same way as the human eye does. With two photographic images taken from different horizontal angles, depth information can be calculated, using a computer, by corresponding the pixels in the left and right images. This is called solving the correspondence problem in the field of computer vision.

The Correspondence Problem

> "Given two or more images of the same 3D scene, taken from different points of view, the correspondence problem is to find a set of points in one image which can be identified as the same points in another image. In computer vision the correspondence problem is studied for the case when a computer should solve it automatically with only the images as input. Once the correspondence problem has been solved, resulting in a set of image points which are in correspondence, other methods can be applied to this set to reconstruct the position of the corresponding 3D points in the scene."[12]

In the same way, humans are able to perceive depth from corresponding points in the two images on the retinas of our eyes, which are positioned slightly apart from one another. Years of research have shown that determining stereo correspondences by computer is a difficult problem[9]. There is no definitive solution and often the accuracy of the results is circumstantial. Current methods appear to be more successful when there are very strong similarities between the two images. This is usually obtained by keeping the distance between the cameras (referred to as the baseline) relatively tight compared to the objects in the environment. If the baseline becomes too large perspective disparities and foreshortening on the objects makes their appearance different in both photographs[9]. Also, occlusion can occur when sections of an object visible in one photograph are obscured in the other, therefore removing potential corresponding points. All these things can make it difficult for the computer to determine accurate stereo correspondences. However, on the other hand, if the baseline is scaled too small, then computing the depth becomes very sensitive to noise in image measurements[9]. Often camera calibration is used to try to control these known problems.

Camera Calibration

By ensuring similarities between the two photographs, as well as optimising aspects such as the baseline between cameras, recovering 3D structure from two photographs becomes a simpler problem. Camera calibration is the term given when the “mapping between image coordinates and directions relative to each camera is known. This mapping is determined by, among other parameters, the camera’s focal length and its pattern of radial distortion.[9]” Camera calibration involves ensuring that the algorithms designed to solve the correspondence problem have an easier job, with more precise results, as certain variables or useful information is made known. Whilst there are algorithms and techniques that are able to compute depth from uncalibrated views, “camera calibration is a straightforward process that considerably simplifies the problem.[9]”

Structure From Motion

> “In computer vision, structure from motion refers to the process of building a 3D model from video of a moving rigid object. Algorithmically, this is very similar to stereo vision where a 3D model is built from 2 simultaneous images of the same object. In both cases, multiple images are taken of the same object and corresponding features are used to compute 3D locations.”[13]

A similar alternative to Stereopsis is Structure from motion. Like Stereopsis, structure from motion techniques use the correspondence problem to calculate depth. The main difference being how the multiple photographs of the object required are obtained. Whilst stereopsis uses stereo vision to capture two images of the subject from different perspectives at the same time, structure from motion relies on camera or object movement to obtain similar resulting photographs, only each photograph will be captured at a different moment in time. This of course has limitations. For example the subject has to be rigid so that pixels in one photograph correspond to pixels of the same subject in another. This is required to solve the correspondence problem. Also, depending on the type of camera movement, it can be more difficult to calibrate the camera.

4.1.3 Practical Application.

Using image-based modeling as a technique to obtain depth information will require additional forward planning and therefore time on top the current filming process. Decisions will need to be made as to what information and extra equipment, other than the standard camera kit, will be required. It would be possible to simply attach another matching camera, either slightly to the left or right of the camera used to film the actual shot. This may or may not suffice in acquiring useful data, but at least this then gives the opportunity to attempt some form of stereopsis at a later stage. In practice, the techniques discussed can not guarantee success or useful depth information every time, however in the circumstances that it does work, the depth information retrieved could be very valuable.

Camera calibration would require any crew or staff on a shoot to have extra expertise and experience in order to combat the potential problems such as noise, occlusion and perspective disparities to achieve effective results. This, along with the additional time, the unreliability of the process and the cost of resources such as software, equipment and staff, means that a decision needs to be made about how valuable is the depth information is against the the potential problems and costs that come with obtaining it.

4.1.4 Advantages.

> “Many real-world objects, such as trees or people, have complex shapes that cannot easily be described by the polygonal representations commonly used in computer graphics. Image-based representations, which use photographs as a starting point, are becoming increasingly popular because they allow users to explore objects and scenes captured from the real world.”[8]

Stereopsis in general, has greater success in urban or built up environments where objects in the scene have more visual cues for perspective. Most architectural constructs strongly resemble an arrangement of basic geometric primitives. This makes it easier to understand what is happening in the scene and therefore calculate depth. Also there is likely to be less visual noise in an urban environment compared to say a forest of trees and leaves. This helps to obtain corresponding pixels and patterns in both photographs.

The process of capturing required data for stereopsis is not very different from ordinary filming. Only the slight change of an additional camera or cameras are required. It would therefore not be a problematic task to modify the current situation to include stereoscopic capabilities.

Creating an image-based model, of an environment can have more advantages than capturing depth information alone. Additional 3D data can be extracted. It would be possible to render normal maps to allow new 3D lighting adjustments to be made to the scene after it has been filmed. It also opens up possibilities for set extension or set replacement with a manipulated fully computer generated version. This would be useful if drastic alterations to the environment, or circumstances that are unavailable, or impossible to film, were necessary to achieve a desired result.

4.1.5 Disadvantages.

Currently there is no method or algorithm that solves stereopsis and the correspondence problem with complete accuracy. As a consequence, there are many different algorithms have been written and developed by different individuals and companies to try and improve on current methods and find the best solution.

> “These systems are only as strong as the underlying stereo algorithms. This has caused problems because state-of-the-art stereo algorithms have a number of significant weaknesses.”[9]

Some of the weakness include problems that were discussed earlier such as occlusion, perspective disparities and foreshortening. All of which cause problems when trying to examine stereo correspondences. Even with camera calibration, there are still many potential problems that can not always be solved, leading to inaccurate results. This makes the whole process relatively unreliable. An understanding of the weaknesses of this system, means that steps can be taken to lessen the impact of the problems. Nevertheless, in a commercial situation, it is not appropriate to utilise an unreliable system and invest in more equipment, software and expertise for something that may or may not provide adequate results.

The location and type of the environment that depth is required for can cause problems when the stereo photographs used to calculate depth are overcomplicated and noisy. This makes calculating stereo correspondences between the two photographs difficult as there are no clear points to follow. Stereopsis can be limited as it tends to favor more simple urban environments with architectural, geometric shapes where it is easier to find continuity between the two photographs, rather than complex organic environments with heavy image noise.

The additional time and cost that could potentially be required to obtain any form of usable data may not be worth what the results can provide.

4.2 Light Detection and Ranging

4.2.1 What is LIDAR?

> "LIDAR (Light Detection and Ranging; or Laser Imaging Detection and Ranging) is an optical remote sensing technology which measures properties of scattered light to find range and/or other information of a distant target."[1]

LIDAR works in a similar way to radar technology, which uses radio waves instead of light. It is widely used in geographical sciences to gather useful data about landmarks, terrain relief and other geographical features. Developments in the technology has meant that LIDAR has many varying applications and uses. Typical applications are found in the fields of Geology, Seismology, Atmospheric, Physics, Remote Sensing, Engineering Geology[1]. In 3D graphics, it can be used to scan real world objects and return a collection of scattered points in three dimensional space (often referred to as a point cloud). These points are formed to resemble the scanned target. The point cloud data can then be used to extrapolate a usable 3D model or depth map image.

> “The primary difference between LIDAR and radar is that with LIDAR, much shorter wavelengths of the electromagnetic spectrum are used, typically in the ultraviolet, visible, or near infrared.”[1]

LIDAR technology scatters pulses of laser light at its target. These pulses of light hit the target and are reflected back. The time delay between the transmission of the pulse and the detection of reflected rays is used then to calculate the distance that the pulse traveled before being reflected back. This is the distance between the scanner and the target. If enough pulses are fired, then it is possible to generate a point cloud that resembles the subject.

Figure 4.1: **LIDAR point cloud image.**[15]
point cloud generated from a LIDAR scan of the A14 to be used by collision investigators after an accident.

Figure 4.1 is a rendered image of a 3D point cloud that was generated using LIDAR technology. This image was scanned from the top of a van traveling at 50mph and is a still frame from a real-time sequence. Point clouds can then be used to generate depth maps or full 3D models.

Figure 4.2 shows the depth map of scenes along a section of street calculated from LIDAR data[16]. Depth maps are simple to create from LIDAR data, as they are essentially the same thing. It is also possible to use LIDAR data to reconstruct surfaces in 3D.

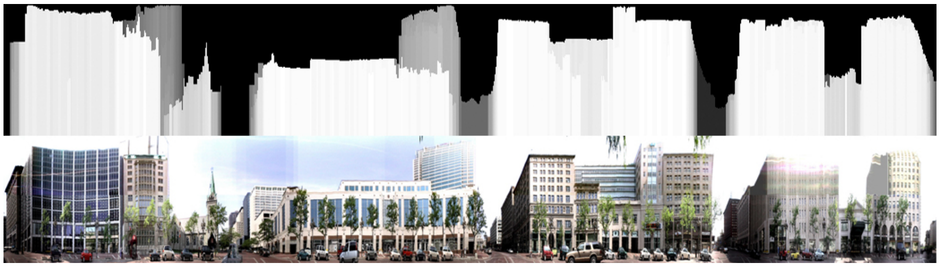


Figure 4.2: **Depth map of buildings from a camera path is predicted from LIDAR data.**[16]

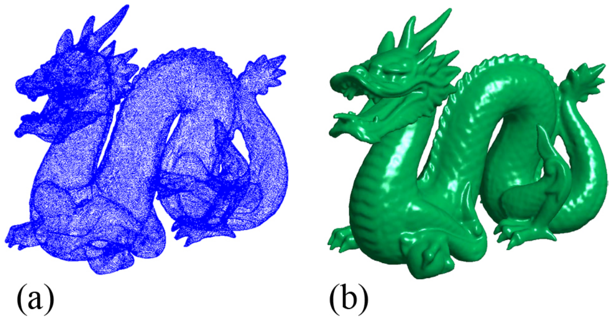


Figure 4.3: **Fitting a Radial Basis Function (RBF) to a 438,000 point-cloud.**[18]

a) A render of the 3D point cloud.

b) Reconstructed Surface.

In Figure 4.3 you can see an example of a reconstructed surface from a point cloud generated with LIDAR data. The surface in Figure 4.3a was created from the point cloud data automatically using a Radial Basis Function (RBF). Another common approach to surface reconstruction is to import the point cloud into a 3D modeling package. The points can then be used as a reference or modeling template, from which a hand modeled polygon replica can be made.

4.2.2 Practical Application.

LIDAR is a versatile technology that is steadily discovering multiple new uses all the time. However, as it is still relatively new, equipment can sometimes be costly. However the results are far more reliable than any image-based depth calculation. As time moves on, this will only improve. Already LIDAR has lent its services to big budget visual effects films such as 'The Day After Tomorrow'. This film made use of LIDAR scanning for providing modeling templates of New York City[7].

Figure 4.4: **A still frame taken from the film "The Day After Tomorrow"[7].** *This shot shows 3D models of buildings that were scanned using LIDAR technology in New York.*

In Figure 4.4 you can see the LIDAR scanned buildings of New York City in the background. The depth information acquired was essential for the success of this scene, where a large tidal wave sweeps through all the buildings. It was required to calculate the interactive fluid effects of the waves movement through the environment. Depth information was also used to help composite the 3D effects in the scene, as well as incorporate 2D atmospheric effects.

The depth information from LIDAR scans is not perfect, but it is reliable and accurate. The results and outputs are predictable, which from a commercial point of view is appealing. It has also proved its worth through commercial success in films like The Day After Tomorrow.

4.2.3 Advantages.

LIDAR provides stable, reliable and predictable results. “The nominal accuracy for a LIDAR scan is 0.006m”[17]. Even though it is not perfect, the errors and problems are not overly challenging to combat. Also, as these errors are understood and expected, tools and methods to designed compensate for them can be easily put into place.

As the popularity of LIDAR increases and demand increases, the price of acquiring this technology will go down. Despite minor errors, the quality levels of and accuracy of LIDAR are high meaning that LIDAR can capture much finer detail than alternative methods of acquiring depth information. Over time, as the technology is developed further its sophistication will increase still. This finer detail can be used to create complex and accurate models which can be extremely useful. Having a 3D model of an environment makes extracting extra 3D information on top of depth possible. As LIDAR is also able to scan great distances, it can gather lots of useful information about a scene or environment. This could save time by replacing the need to locate plans and take measurements of a set.

4.2.4 Disadvantages.

> “LIDAR data is often noisy, irregular, incomplete and redundant, due to:

- limited range resolution.
- misregistration between scans taken at different scanner positions.
- restricted physical viewing angles for the scanner.
- increased susceptibility to occlusion.
- over-lapping scans.” [17]

Because of these problems, LIDAR data will often need processing and filtering after a scan in order to achieve acceptable results. Sometimes the data is too messy to expect any software to do a good job of automatically cleaning the data. In these circumstances it would be necessary to use the point cloud data as a model template to produce a 3D model by hand. If this is the case, the usefulness of the data is reduced. Automatic or manual data cleanup takes more time, money and resources. As a tool, it has not been developed specifically to capture depth information that compliments live action footage. To capture LIDAR data will require additional equipment, most of which will be costly.

Chapter 5

Live-Action Depth Map Examples

To demonstrate some potential uses that depth maps could provide for live-action footage, I have created a basic depth map for a photograph (Figure 5.1a) from a simple 3D model that I created to resemble the scene. I then performed some simple crude tests using techniques previously described in section 3.1 Examples of 2D Uses For Depth Maps. This is to demonstrate the potential usefulness a proper depth map image could have.

Using the depth maps in Figures 5.1b and 5.2b, I was able to use Z Depth Compositing to integrate the CG with the live-action. The results that can be seen in figure 5.3 are only let down by the inaccuracy of the depth map created from some very basic dummy geometry.

I also performed some rough depth of field defocus effects (Figure 5.4b) , as well some simple fog and atmospheric effects (Figure 5.4a). The effects are very obvious to illustrate a point. With a little more time and care, it would be possible to produce more convincing results.

As I created a basic 3D model to acquire some a fake depth map, it was possible to obtain the normal information of the model also (Figure5.5b). I finally attempted to use the NormalLight3D[5] macro in shake to add virtual 3D light to the scene (Figure 5.5a).

The relighting in Figure 5.5a was only very basic and so therefore the results look crude and far from real. The most obvious effect is the specular highlights on the trees in the centre of the photo. With a little more time, care and proper normal and depth data, this could be an extremely useful tool for refining a live-action plate.

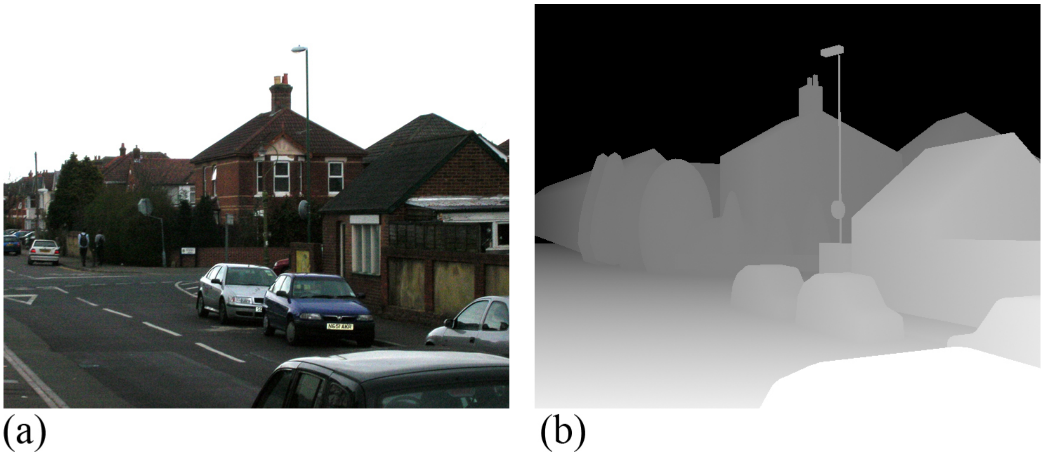

(a) (b)

Figure 5.1: **Still Photograph with Depth Map image.**
a) Photograph of a simple suburban environment to perform tests on.
b) This is a depth map image for the photograph. This was created for test purposes. It is very basic and was made by modeling primitive shapes to resemble key objects in the photograph.

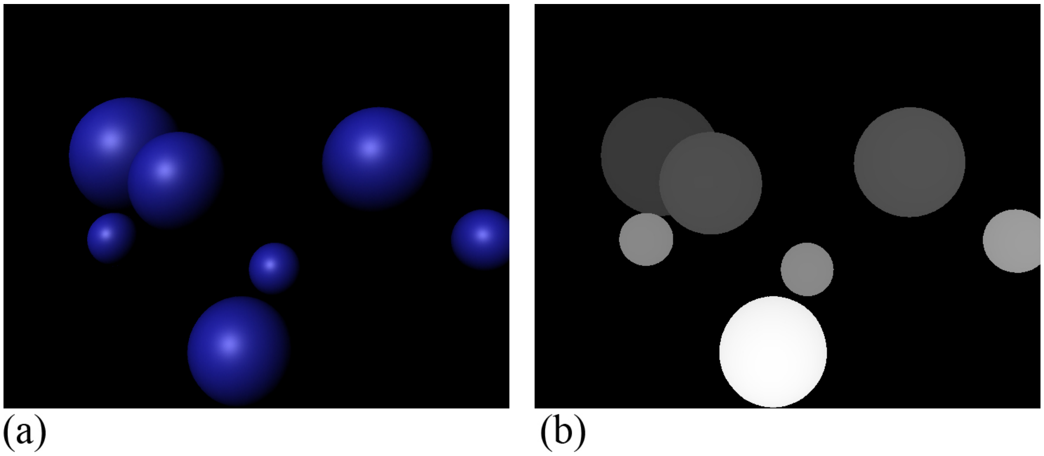


Figure 5.2: **Render of 3D scene with Depth Map to be Z Depth Composited with the Figure 5.1a.**

a) Rendered image of a 3D scene.

b) Depth map image for Figure 5.2a.

Figure 5.3: **Z Depth Composite of Figures 5.1a and 5.2a**


(a) (b)

Figure 5.4: **Depth map used to add effects to Figure 5.1a.**
a) Atmospheric fog effect added to photograph in Figure 5.1a.
b) Depth of Field Defocus effect added to photograph in Figure 5.1a.

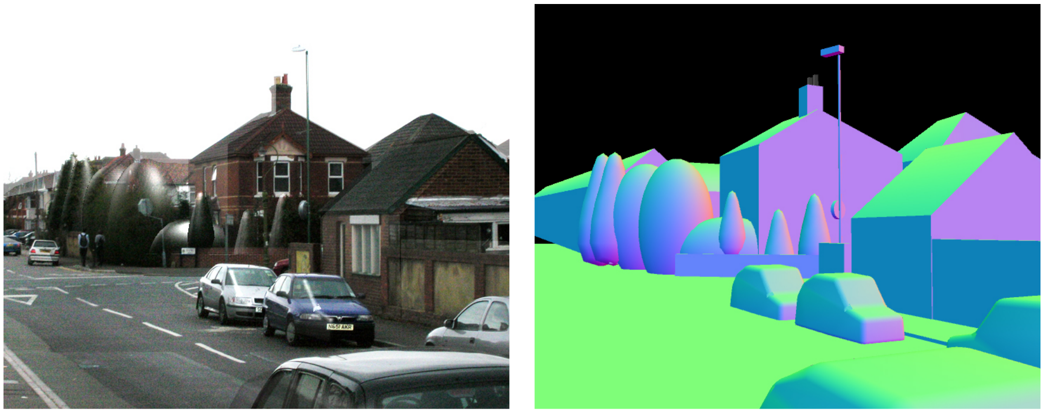

(a) (b)

Figure 5.5: **3D lights added to Figure 5.1a in shake using NormalLight3D[5] macro.**
a) The photograph in Figure 5.1a has rough 3D lighting added to it in 2D compositing package Shake using a depth map (Figure 5.1b) and a normal pass (Figure 5.5b).
b) The normal pass rendered from the dummy geometry created for the photograph in Figure 5.1a. Used to relight the scene.

Chapter 6

Report Conclusion

6.1 Are Depth Maps For Live-Action Useful?

I believe that the answer to this question is yes. The mini tests that I performed in Chapter 5 would have taken more time and effort if not proven impossible to do without a depth map image available. There are many more uses and purposes for depth maps that have not been included in this report. In order to achieve a precise integration between computer generated imagery and live-action footage, I feel that it is important to have as much information about the scene as possible at the artists disposal.

When combining computer generated imagery and live-action footage together, having depth information for only one of these seems wasteful, as it is not very useful on its own. However if every element in the scene has depth information, then the process of determining occlusion can be automated. This is a big step to combining both elements together, and because it is an automatic process, it is extremely fast. This means that more time can be spent focusing on other important integration problems.

6.2 Are Depth Maps For Live-Action Worth the Cost or Potential Problems to Get?

I believe that this depends on the shot. There is no point spending large sums of money and time to receive a depth map that there is no need for. In more simple scenes, depth maps do not have such a high advantage over manually masking and keying required elements. In a situation where you have lots of CG objects occluding live-action objects and/or vice-versa, depth information would be incredibly useful to work out what needs to be seen on top.

There is evidence that they are worth while by the fact that big visual effects films are already using them. Take the example of The Day After Tomorrow again. They achieved spectacular visuals that may of not been possible without depth information gathered from LIDAR scans. As visual effects become more and more complex, acquiring information like Z depth data will become more and more important.

6.3 How Successful was this Project?

Through the process of researching and writing this report, I have learned much about 2D compositing and the methods employed to create and layer up elements of a shot. I realise how the spacial relationship between objects in a scene is extremely important for achieving a precise integration. I now feel that I have a deeper understanding about the work flow and bridge between 3D computer graphics and 2D computer graphics. The knowledge that I have acquired will be extremely beneficial to my major project and will have a strong positive influence on the techniques that I exercise when producing the shots. This has provided me with a stepping stone into 2D computer graphics, something that I have previously been unfamiliar with.

I am glad that I chose to make this project a report based project. It has given me time to read and think about the a broad scope of concepts that I wanted to learn. It has also been good writing practice for my major project report. Whilst I would have also enjoyed creating some form of program or having some kind of product outcome from this project, I also know that there will be many opportunities to implement my ideas in the future.

Given more time or a repeat chance. I would take the opportunity to explore other bridges between 2D and 3D graphics. Depth Maps are a small part of massive and interesting subject.

Bibliography

[1] BRINKMANN, RON., 1999. The Art and Science of Digital Compositing.

[2] SANTIAGO, DAVID., 2005. Creating 3D Effects for Film, TV and Games.

[3] BIRN, JEREMY., 2000. [digital] Lighting & Rendering.

[4] PAOLINI, MARCO., 2004. Shake 3 Professional Compositing and Special Effects.

[5] DRESSEL, DERICK., 2006. A macro download for Shake called NormalLight3D 1.0.0 [online shake macro]. http://www.highend3d.com/shake/downloads/macros/filters_effects/3810.html.

[6] WRIGHT, STEVE., 2002. Digital Compositing for Film and Video.

[7] TWENTIETH CENTURY FOX, 2004. The Day After Tomorrow. Bonus Features Disk [DVD video].

[8] B. MOK OH, M. CHEN, J. DORSEY, F. DURAND, 2001. Image-Based Modeling and Photo Editing. Proceedings of the 28th annual conference on Computer graphics and interactive techniques SIGGRAPH '01. Publisher: ACM Press.

[9] DEBEVEC. P, TAYLOR. C, MALIK. J, 1996. Modeling and Rendering Architecture from Photographs: A hybrid geometry- and image-based approach. Proceedings of the 23rd annual conference on Computer graphics and interactive techniques SIGGRAPH '96. Publisher: ACM Press

[10] WIKIPEDIA, 2006, Stereoscopy [online]. Available from: http://en.wikipedia.org/wiki/Stereoscopy [Accessed 6 March 2007].

[11] WIKIPEDIA, 2006. Stereopsis [online]. Available from: http://en.wikipedia.org/wiki/Stereopsis [Accessed 6 March 2007].

[12] WIKIPEDIA, 2006. Correspondence problem [online]. Available from: http://en.wikipedia.org/wiki/Correspondence_problem [Accessed 6 March 2007].

[13] WIKIPEDIA, 2006. Structure From Motion [online]. Available from: http://en.wikipedia.org/wiki/Structure_from_motion [Accessed 6 March 2007].

[1] WIKIPEDIA, 2006, LIDAR [online]. Available from: http://en.wikipedia.org/wiki/Lidar [Accessed 6 March 2007].

[15] STREET MAPPER, 2007. Mobile Mapping Using LiDAR Technology [online]. Bingham, Nottingham. Available from: www.streetmapper.net/casestudies/A14.htm [Accessed 6 March 2007].

[16] J. Y. ZHENG, X. WANG, 2005. Pervasive Views: Area Exploration and Guidance Using Extended Image Media. Publisher: ACM Press.

[17] APPLIED RESEARCH ASSOCIATES NZ LTD, 2007. Reconstructing surfaces from LIDAR data [online]. Available from: http://www.aranz.com/research/modelling/surface/applications/lidar.html [Accessed 6 March 2007].

[18] J. C. CARR, R. K. BEATSON, J. B. CHERRIE, T. J. MITCHELL, W. R. FRIGHT, B.C. MCCALLUM, T. R. EVANS, 2001. Reconstruction and Representation of 3D Objects with Radial Basisl Compositing.

[2] SANTIAGO, DAVID., 2005. Creating 3D Effects for Film, TV and Games.

[3] BIRN, JEREMY., 2000. [digital] Lighting & Rendering.

[4] PAOLINI, MARCO., 2004. Shake 3 Professional Compositing and Special Effects.

[5] DRESSEL, DERICK., 2006. A macro download for Shake called NormalLight3D 1.0.0 [online shake macro]. http://www.highend3d.com/shake/downloads/macros/filters_effects/3810.html.

[6] WRIGHT, STEVE., 2002. Digital Compositing for Film and Vi Functions. Publisher: ACM Press.