Implementing virtual sets in a 2D environment

A study in how to efficiently integrate 3D virtual sets with photos

David Nilsson
Abstract

In this paper I’ll be investigating how to implement and use virtual 3D sets using only simple 2D reference footage. There are many possible ways to achieve similar effects but I’ll be specifically targeting the lower end segment of the industry by using methods that require fewer resources. Other methods are used in large film productions in Hollywood but my goal with this paper is to make the technology more accessible for small-scale companies and productions.

1. Introduction

A virtual set can be an extremely useful tool and the potential fields of application are unlimited. Today, the usage of this technology is extremely confined and its by far greatest employment is in multi-million dollar Hollywood movie productions. With this paper, my aim is to make the application more accessible to small-scale productions and to alter the perception that such techniques can only be achieved by large studios.

In movies, the large studios generally use video footage that they’ve shot themselves; if they require a shot with the Eiffel Tower, they fly an entire camera crew to Paris, possibly rent a helicopter and close the streets to be able to shoot the 20 second clip they need. These expenses compounded with equipment costs add up to a considerable sum and are, naturally, not feasible for small-scale productions.

What I’ve therefore concentrated on is how to achieve similar effects with very limited budgets and resources. Using the methods explained in this paper anyone with a basic knowledge of 3D modeling can create virtual sets using only a single photo and a 3D application. As these are the only requirements, it’s also fairly simple to find the necessary material for free on the Internet by browsing large photo websites such as Flickr.

To mention one example, this technology could for instance be used in distance education where an educator with limited resources could appear to be teaching from a modern office rather than in his one bedroom apartment.

Throughout this paper, I will guide you through the process using my own example combining a shot of central Umeå with a 3D model of Stonehenge, inserting Stonehenge into the town’s central square. To make the video clip look “alive” and real (as opposed to a still photo), I will also add some camera movement to the shot.
2. Related Research

The technology explained in this paper can be greatly expanded upon. For instance, it’s possible to record footage with a “green screen” with which you can quite easily add humans and other moving objects to the scene and even have them interact with the environment, although interaction with the environment should be done with caution as it complicates the shot, requiring interaction with the 3D set.

A combination of the technologies explained in this paper, that is, a virtual set based on a still image and adding moving objects and people to that scene using green screen footage is being used in movie and TV series productions around the world. One example, as mentioned on the fxguide website, is the British sci-fi series “Red Dwarf” that has used the technology for a range of different shots. One example being where they took a picture of central London and convincingly inserted an alien pyramid structure into the city. To make the scene look more realistic and alive, they also added camera movement in conjunction with other moving objects such as the Red Dwarf “carbug”, people (recorded on green screen) and effects such as lighting to the shot, making it look very convincing and not at all low-budget.

It’s also possible to use “matchmoving” to improve the shots and make them feel more realistic. This can be used on different levels. One thing matchmoving can be used for is the camera, you can record the way you want the camera to act in your shot and import the exact movement into your scene generating a more realistic movement. To mention an example used by Robert, T (2005), imagine that a CG creature is going to crash out of a window and run across the street into an alley. If the director has a good idea of how he wants the camera to move during that shot, he (or his cameraman) can record the camera movement he’s looking for and then import it into a matchmoving software after which it can be exported as a camera that can be used for the shot in a 3D application. It can be quite complicated to create realistic camera movements in 3D applications, so importing a “real” camera movement can be very helpful to perfect the shot.

Taking matchmoving to the next step, we’d be leaving the still image technology behind, instead using video footage for the environment and placing the 3D objects directly into the video footage. This way, you won’t have the problems that you would when using the technology with a still photo, which I will expand upon in this paper. For instance, you don’t need to model your environment and obviously camera movements won’t be an issue as texture information won’t be missing when you reveal new parts of buildings.

This particular field of research is still relatively new and there are different kinds of research that can be done. While I’ve focused on explaining the practical process of applying the virtual set technology with still images, a whole different approach has been taken by Lourakis, M. & Argyros, A. (2005) who have written a very technical and theoretical paper focused on how to use camera tracking with matchmoving technology in complex environments.

As concluded by Robert, T. (2005), the technology that I’m working with, using still images to create virtual sets works well for relatively simple still image situations, but for more complicated shots, it is useful to enlist the help of matchmoving software such as
“Boujou”. This is more similar to the workflow used by the major film studios and it requires video footage instead of still images.

3. Problem & Purpose

The incentive behind my project was to be able to answer the question; “How can people and companies with modest budgets create and use virtual sets in their productions?” Hence, the goal with this paper is to make 3D virtual set technology accessible for productions made by these markets. If I succeed, I hope to see this method used in many different fields, creating greater possibilities for these groups.

To name a few potential areas of use, it could be tremendously helpful for educational purposes; one example being distance learning. Small-scale companies might not have the resources or locations to make a serious impression on potential clients and students, in this case they would be able to resolve the issue by creating a good-looking virtual classroom or office. This environmental flexibility could also be applied to the education programmes themselves; a history teacher, for example, could appear to be walking in prehistoric surroundings while discussing the era, making learning much more interesting and enjoyable.

Closer to home the technique could be utilised during the festivities for the “European Capital of Culture” 2014 hosted here in Umeå. One of the big concepts is to show Europe from Umeå’s perspective and using these methods the planning committee could easily and economically create informative and interesting productions presenting this theme. As one would expect, these methods would also prove highly useful in movie productions, advertisements and so on, allowing increased creativity thanks to its affordability.

4. Realization

4.1 Method

There are several possible ways to approach the aforementioned problem, for example to try to answer the question purely theoretically by researching literature and drawing conclusions from it or to approach people and companies currently using similar technologies to ask them about their experiences, take notes and present that information and theoretical solutions.

I chose to approach the problem in a more practical manner by creating a prototype and in the process a clear method or workflow that I know can be applied by others looking to utilise these effects. This technology is to a very large extent about handicraft and therefore I want to be able to demonstrate for people how to use these effects themselves rather than having them rely on a theoretical paper and conclusions drawn solely from topical literature. This clarity is especially important since my main goal is to make the technology more accessible.

Through my own experience creating the prototype I will gain greater insight into the technology and experience first-hand any possible hindrances to the workflow, hopefully gaining the knowledge required to explain fully and coherently how to overcome them as well as defining the optimal workflow. Using one of the other approaches I would have had a
much more distanced result and would most likely not be able to explain how to approach the project as thoroughly.

4.2 Planning
To begin with, I considered my options and ideas for an example video with which I would be able to show the procedure and results using my suggested methods. I needed to pick an environment that would be spacious enough to house any object I chose to implant there and at the same time prove sufficiently challenging to highlight issues that might arise with the method.

I eventually decided to take a photo of the centre of Umeå hoping to provoke potential obstacles to my theory by using a space that would naturally be filled with objects and insert Stonehenge into the central square. My reasons for choosing Stonehenge were its worldwide recognition, due to which it would be clear that the landmark had been artificially introduced and its simple construction to avoid absorbing too much project time in the modeling process, a subject which is covered thoroughly by other literature.

I was originally considering creating a more advanced clip including an actor whom I had planned to have interact with Stonehenge using an interactive user interface, for instance he could possibly move the stones around, or he could simply get more information about Stonehenge on an interactive information screen. However, after careful consideration I decided to stick as much as possible to the basics to increase accessibility for smaller scale productions as that was the main goal with this project. I will however discuss how to expand the technology further later in this paper.

4.3 Workflow
Commencing the project, I photographed the environment that I was going to use for my shot. I chose to take the shots on a Sunday in order to get a photo of Umeå’s central square with as few people as possible. The reason behind this being that in the final clip it would quickly become apparent that it was in fact only a photo and not video footage as the figures would be standing still throughout the shot.

Figure 1. Original photo footage.
My original photo (see Fig. 1), however well planned, still needed revision as it contained a dog and a number of people that I was forced to edit out using Photoshop. As I wanted to insert Stonehenge directly into the middle of the square, I also needed to remove some of the features of the square itself, for example most of the statue taking up a considerable part of its eventual position as well as the lamp post with the bright yellow sign and one of the large columns on the left-hand side to make sure that they wouldn’t disturb the final composition.

![Edited photo](image)

*Figure 2. Edited photo*

The final edited version (see Fig. 2) was by no means perfect, there were some blurry sections and slivers of statue still visible, but I deemed them trivial since those flaws would eventually be concealed once Stonehenge was inserted into the image.

The next step was to create a 3D model of Stonehenge itself. I used a range of reference footage (see Fig. 3) covering different angles of the structure and made a somewhat simple copy in Maya. My decision to keep the model simple, I hoped, would keep the focus of the paper on the development of the virtual set rather than diverting focus to the 3D modeling process.

![Reference footage from Stonehenge](image)

*Figure 3. Reference footage from Stonehenge*
Once the model was finished (see Fig. 4) it was time to start composing the shot in Maya. First of all I inserted the edited photo as an image plane, meaning it served as a background. As I was going to add camera movement to the shot and it wouldn’t look good or believable to pan, zoom and rotate in a still photo, I had to create simple models of the buildings and the obstacles in the shot and align them in their correct positions using the image plane as a template (see Fig. 5). I chose to create basic shapes for the buildings but if the shot would be for a large production, I’d recommend doing a more detailed job here as it can improve the sense of depth.

Once the buildings and obstacles were blocked out, I assigned a so-called “surface shader” material to the objects in the scene and set it to projection mode - this means that you can project a texture onto the objects by using a photo. A surface shader ignores all lighting and shadow information in the scene and instead uses the colours available in the chosen photo. For the projection in this case, I therefore chose the edited photo from the city square. Doing so, all the buildings and objects I had modelled immediately took on their correct textures.

At this time, I had to go into the perspective mode and work a bit on the overview of the scene to make sure that the relative distances between the buildings and objects in the scene were correct. During the modeling process I had mostly worked using the first-hand camera viewpoint. Initially it didn’t occur to me how unusual the depth perception would be and how important the objects’ locations would be, I worked on the image simply overlaying the image with models and the results were far from satisfactory. For instance, even though from the camera viewpoint it looked completely normal, the two buildings on the right hand side were almost touching and if left that way there wouldn’t have been enough room to insert Stonehenge into the square. Even some smaller details were deceptive, for example the far most lamp post seen from the overview was actually closer to the camera than that on the left. Looking back at my approach it seems obvious that my tactic was flawed.
When I was satisfied with the overview I imported my Stonehenge model into the scene and aligned it in the centre of the square (see Fig. 6). At this point, I was more or less done with the still scene. Going back to the original camera view and performing a test render at this time gave me the result shown in Figure 7.

Specifically note the details in this render that show us that this in fact is a 3D scene. For instance, it’s clear that the bin and the lamp post in the foreground are closer to the camera than Stonehenge. Of course, this could potentially just be advanced image editing, and that’s why we’re going to add camera movement to the shot to make it more believable. To do so, start out by creating an additional camera and animate it until you’re pleased with the movement.

The camera movement was by far the most advanced problem. Because we’re using a 2D photo to project the textures onto the buildings, when adding camera movement to the 3D version new parts of the scene are unveiled that technically didn’t exist in the original and therefore lack texture information. Also, the background image still contains all objects, it’s the same image as that used for the texture projection. This means that if you were to rotate the camera a little, you would for instance be able to see not only the foreground lamp post whose texture is projected onto the 3D model but also its 2D image, still present in the background image “behind” the foreground model.

This problem resulted in my having to create several different editions of the projection image as well as an updated background image. I was obliged to create a clean backplate in
Photoshop where I replaced the buildings with sky using the clone stamp to avoid doubled architecture. I also needed to create a projection image for the ground, in which the lamp posts and the remaining obstacles there were removed and replaced with paving. Finally I had to create some specific projection images for the sections of buildings that lacked texture information, for instance the main building on the left was initially covered by a birch tree and columns belonging to the outside seating area (see Fig. 2) so I had to create a projection image that filled in the blanks.

When I made a test render of the scene with the camera movement added, I noticed some additional problems to be addressed. My first concern was with the reflections of scenery in the windows on the right-hand side. As the camera was moving, the window reflections remained consistent, reducing the scene’s plausibility. I therefore had to edit out the window reflections in the source image to make the shot more believable. Finally, the scene suffered from a slightly lifeless feeling and I tried to resolve this by creating an additional light source so that light would fall more naturally onto Stonehenge and help it blend into the realistically lit projected image. Unfortunately without manipulating the light to an unrealistic degree the artificial ‘sun’ fell on a limited section of the structure and I was forced to resort to more complicated means in order to give a more convincing look to the stones (see Fig. 8).

Because surface shaders ignore shadows and light in the scene and I wanted the stones to be able to cast shadows on the ground, I had to find a workaround for that problem. After some research, I found help on CGSociety where a member had created a surface shader with this feature that I could use for my scene. I would however say that I believe that the effect would have been more convincing if I had taken the photo at a different time of the day which would have resulted in more shadows being casted.

![Figure 8. Final still render](image)

Once this was done, I rendered the 125 frames of animation and imported them into a video editor where I finally exported it as a video clip.
5. Results and performance

As mentioned previously, one of my goals was to investigate the accessibility of Hollywood-like effects for low budget productions. This test illustrates that it's possible to achieve Hollywood-like effects with next to no external resources required. All it takes is basic knowledge in 3D software and time: depending on how complex the environment that you choose is, it can also require knowledge in photo editing.

I feel that I perhaps should have spent some more time texturing the stones as I believe the video clip isn't fully convincing in large part due to the unrealistic quality of the textures and that it would also have added to that effect if the area around the stones exhibited greater signs of wear or perhaps had a little moss growing on both them and the ground to imply permanence in the location and to blend them further into the surrounding area. However, as I mentioned earlier in the paper, my purpose was to investigate the potential use of these kinds of virtual sets for small productions and review the workflow one should use to implement them.

I chose a complex environment for my test on purpose to be able to illustrate many problems that one might run into when applying this technology and in all honesty this method hasn't proved effective with complex environments. There were many objects in my original photo that caused problems; trees and lots of small objects such as bikes in the background were only some of them. Also, it would have been easier if I had taken a photo from farther away as less “unknown” (to the camera) objects would have been revealed by a more subtle camera movement and problematic areas wouldn't have been as readily noticeable.
I'd recommend that hobbyists try to stick to more simple environments – inserting Stonehenge on a meadow in the neighbourhood for example would be a pretty simple task due to the limited requirement in terms of modeling. Creating convincing shadow effects would also be simpler as they would be more visible on a model inserted into an environment with a greater amount of open space. However, with a lot of time on your hands any environment (including the one I chose) could be made very realistic. One should also bear in mind that camera movements make the procedure much more complicated so the best use of this technology is for short clips. If you’re planning on moving the camera a great deal, it’s a good idea to acquire a large quantity of photo reference material from the area so you can add original textures to the surrounding objects that will be revealed by the movement in the scene.

6. Conclusions

I think that the technology discussed in this paper could prove useful for many different kinds of productions. As mentioned previously, one area where I could see this basic virtual set technology being implemented would be in distance learning, which is becoming a more and more common way to study with the Internet development. Software developers Adobe (creators of popular software suites Photoshop and After Effects) have already created virtual classrooms and offices using similar methods for their Photoshop tutorial series shown on their “Adobe TV” website.

Another good example of use would be in architecture. It would be very convenient for architects, estate agents and city planners to be able to easily place the model of a structure on a plot to see what it would look like in their planned environment. Some other potential areas of use are teaching, commercials, music videos, small-scale TV and movie productions.

The technology presented offers great freedom as one doesn’t need to own an expensive video camera: it’s possible to acquire reference footage from extensive image sites such as “Flickr”. Due to the abundance of footage on these sites they provide adequate reference material leaving the production team free to simply build their scene within the chosen environment even if it would usually be too expensive or difficult to access. Using these virtual sets also provides much flexibility to your productions. The fact that you’re able easily exchange objects in the virtual set helps in making the technology powerful.

The goal when employing this method is to create a virtual set that seems realistic. As written in the works of Lucinelma Pessoa Albuquerque, A. et al. (2000): "The final goal of a good virtual set is to attain total integration between real and virtual worlds: a person from the audience should not distinguish one from the other". I agree completely with this and I wouldn’t advice using this technology if you’re planning complex shots with difficult environments and a lot of movement, acting or interaction with the set.

While it’s true that most environments can be made realistic using the methods I’ve explained, complex environments would require a great deal of time to perfect. Instead of having a team work on that environment for a long period of time, it would most likely be more efficient to capture real video footage of the environment instead and implement the virtual objects using so-called “matchmoving” technology instead.
There are advantages and disadvantages with both methods. For instance, while real video footage doesn’t require you to model the environment, it’s much easier to edit the environment to your likings in a single still image than it is to edit the environment in video footage at 25 frames or images per second.
7. References


Web Resources:

Adobe TV with Julieanne Kost:
http://tv.adobe.com

CGSociety: Surface Shader with Shadows:

Red Dwarf: Red on Red & the vfx:
http://www.fxguide.com/article528.html