Joshua Glazer vbunny@mit.edu
Kenneth McCracken kdmc@mit.edu
Alex Sverdlov sverdlov@mit.edu

Man vs. Machine

6.837 -- Introduction to Computer Graphics
Final Project -- team 01
November 30, 1999

Abstract

This paper presents an account of experience gained by implementing a final project for Course 6.837 Introduction to Computer Graphics at MIT. The team consisting of two seniors and a graduate student created a video of a live person performing martial art techniques against a computer generated adversary. The film was shot using a Canon XL-1 digital movie camera, and subsequently edited using Poser 4.0 and Final Cut Pro 1.0. These and other tools were used to render the digital character in the context of filmed scenes, and to fine tune the interactions between the two opponents. It was the group's experience that the significant portion of technical challenges were eventually overcome through skillful use of the tools. However, a few of the project's objectives needed to be met by writing and using C routines to manipulate the decompressed screen shots directly. The finished video was accompanied by a synchronized soundtrack composed by Joshua Glazer, and concluded with a few of the more humorous out-takes from the filming.

Keywords: computer graphics, 6.837, MIT, video, Poser, Final Cut, Tsai method, martial art, modeling, compositing, c code

Introduction

The original idea for this project was conceived by Josh Glazer, with Ken and Alex signing onto the project during the team forming period. It was the consensus among the team members that creating a video of this kind would

- put to use skills gained in the course
- teach us a thing or two about the tools pertinent in the computer graphics industry
- be rewarding from a creative standpoint

Indeed, in the process of making the film we had to draw extensively upon 6.837 course material to determine lighting, shadowing effects, plan camera and lamp positions, as well as to accomplish correct compositing between the digitized character and the filmed scenes. We have also gained experience of researching industry publications that dealt with solving problems similar to our own -- most notably papers by Roger Tsai on camera calibration [5], and on analysis of 3-D time varying scene [6]. We have

learned to use the results presented in these articles as an aid in understanding and meeting our project's challenges. In turn, we hope that the present report of our efforts and mistakes may serve as a point of reference for others in the future.

Goals

Our final goal is a 1.5 - 2 minute digital video depicting a martial arts contest between two opponents. While one of the fighters is human, the other is a modeled animation superimposed onto the scene. The key difficulty of this goal was making the combat appear realistic and engaging, encouraging the suspension of disbelief. The scenes had to be modeled carefully, with attention paid to issues both obvious and subtle. While much time was spent getting such macro effects as shadows to look convincing, many smaller details of animation were also implemented. For instance, Fred's (he had to get a name, since we couldn't really say 'computer-generated-animated-3D-character' hundreds of times on end) mouth moves synchronously with his speech. He also shifts his eyebrows, and blinks at random intervals, adding to the realism of the scene. To accomplish the necessary effects, the main goal was partitioned into the following tasks:

- storyboarding the battle before hand
- shooting the film itself
- importing and editing the footage
- modeling Fred's anatomy, his clothes and weapons
- teaching Fred's body to react realistically to the techniques dealt out by Sensei Godfrey Inniss (human character)
- modeling the interactions of the fighters' weapons and body parts during battle (compositing of the pixels)
- tracking the position of the camera
- compositing Fred with the live video
- creating the shadowing effects of the animated figure in the live scene
- and finally adding an award winning soundtrack

Each of these items was addressed in turn, with the next section of the paper detailing our successes and failures at each of the stages.

Achievements

Storyboarding

Our first material step on the road to this blockbuster was creating a storyboard. We sketched the scenes that introduce the characters, planned the positions and angles of the shots depicting the battle from various points of view, and scripted a quality dialog in the proud tradition of imported kung-fu movies (see appendix H). As a location for the shooting, we chose the seldom trespassed corridor in the basement of Building 9 at MIT. This hallway, with its chain-linked fences and extensive piping provided an original and engaging setting for the conflict.

Our plan was to shoot the live video and to insert the animated character into the backdrop later. Using a stand-in adversary to temporarily cover for Fred would require the eventual removal of one fighter from the video. This approach would risk possible loss of data as per the ray tracing [1] paradigm, in scenes where Fred did not completely cover his stand-in. Consequently, we opted to have

the live character fight a nonexistent foe. Striving for realism, we sought out a consultant (soon turned star) Sensei Godfrey Inniss. He helped us complete the storyboard with specific techniques drawn from his background in Jui-Jitsu. With the plot in place, we decided on the finale, and scheduled the shoot.

Filming and Editing

Each scene in the movie was shot twice. First, with Godfrey practicing a move on one of us acting as an 'ookie' (a recipient of technique), and then executing the same strike or throw against an imaginary opponent. The first version served as a template for animating Fred's movements, while the second take was the actual footage used in the finished film.

While shooting, we were challenged to try to limit the number of sequences in which the camera was moving, while striving for the best possible coverage of the action. This was a concern because we knew that modeling the details of the scene shot by a moving camera was significantly more time consuming. However, many sequences in which Fred appears alone did involve motion of the camera itself in order to give the viewer a more exciting perspective.

To provide light for the filming, we used two bright 300W work lamps. We tried to keep these stationary to simplify modeling. Still, there are more than a few shots in which we were pressed to move one of the lamps because it was producing too much glare. Since the corridor was dully lit, we made the assumption that the hallway's own lights contributed only a muted ambient effect on the scene and needed not to be modeled. Such lighting was thought to have effects approximating to radiosity effects generated by a large, weak area light on the ceiling [2].

In the end, we were satisfied with our footage since only one of the filmed scenes had to be abandoned during editing. This was done because the technique in the scene was an especially involved one, and even after attempting a few takes we did not find them consistent enough to determine (or to realistically model) Fred's position and movements. After trying to remedy the situation by marking the exact positions of Godfrey's feet, and failing yet again, we finally dropped the scene. However, our overall impression was that we did achieve rather good variation between camera angles (experimenting with the shots from the top of the fence, the ladder, tripod panning, and from below the action for the 'headbutt' sequence).

Overall, filming was wrapped up in four hours, totaling 19 minutes of screen time, eventually reducing in size to 1 min 46 sec final version. Meanwhile, the 19 minutes of footage were imported onto a Mac G3 machine via a firewire port, and then DV-NTSC compressed (IEEE1394 standard) with 29.97 frames/sec, millions of colors, and a 720x480 resolution. Each minute of the film occupied ~200MB of disk space and was full dvd-quality. But even so, the movie at this point was still lacking it's main character.

Modeling

To create Fred's basic anatomy we utilized a human skeleton found in Poser's arsenal of stock-made objects [see MetaCreations site [10] for a list and descriptions). The joints of the skeleton could be manipulated via setting angles for bend, side-to-side, and twist motions. Poser attempted to make this task easier by offering inverse kinematics capabilities. Inverse kinematics allows the ability to place the palm and a shoulder in desired locations, leaving it to the software to figure out the data for elbow. However, this 'autofit' feature did not perform to our expectations, as it consistently opted for the most awkward and even physically impossible positions, without hesitation bending human knees the wrong way. In order to achieve a more realistic look and feel, we ended up having to set parameters for all joints, in almost all cases, manually.

Animation of the mouth and eyebrows was achieved through a slightly different technique. In this case, Poser provided an interface to have the lips fold accordingly to five sounds <o, f, t, m, l>. All others had to be obtained through combinations of these, or tweaked manually. A brief slow motion

video of a human reciting film's dialog was shot, and used as an aid in getting this effect.

The trenchcoat and pants worn by Fred were also found among Poser's stored objects. The clothes were spline curved to respond to the character's motion, and only rarely had to be individually positioned from frame to frame. Although Poser did offer the capability to automagically fit the overcoat onto a moving model, this feature too did not work well enough. In a few of the scenes Fred's shin was rendered incorrectly, sometimes showing through his pant leg, and had to be masked out or made invisible on an individual basis.

The tool used to create the sword was Infini-D 4.5, chosen for its solid geometry modeling capabilities. The blade was first prototyped as a flattened polygon, and then extended to match the perceived length of Godfrey's real weapon. While much labor was invested into modeling Fred himself and his accessories, perhaps an equal amount was saved by Poser's ability to do a good job at compositing the character onto the filmed background. Benefiting from this, we were able to concentrate on achieving realistic interactions between the dueling fighters themselves.

One of the challenges in modeling the actual battle was imitating the effect of a sword vs. sword impact, and to a lesser extent of a strike vs. block interaction. In a physical world, the contact is made twice, with objects momentarily 'bouncing away' from each other and coming back together. We were able to simulate this property by using linear extrapolation to marginally speed up Fred's sword immediately prior to impact. This simple technique, coupled with an appropriately timed sound effect, made the difference between a scene 'working' or not.

The antialiasing being handled by Poser, the task of handling occlusion correctly was a less subtle, yet significantly more laborious one. This had to be performed every frame when Fred was found in front of either his opponent or a stationary object. The process itself was straightforward as either all or just those unblocked parts of Fred were superimposed onto the video. Whenever Fred was blocked by Godfrey, Final Cut Pro was used to remove the parts Godfrey covered from Fred before adding Fred to the scene. Whenever Fred was blocked by some stationary object in the scene, that straight-edged object was cut out of the live video and placed atop the combination of Fred and live video, covering the correct part of the digital character.

Camera Positioning and Calibration

In order to assist in modeling the Fred and his background for shadowing effects, we had to approximate the world space position of the camera and lights. The position and orientation of the camera was necessary to better model Fred in the scene, by showing him in the correct size and perspective for proper ray tracing onto the screen. In order to figure out the camera parameters, we placed several bright yellow points in random places throughout the scene we shot the live video from. The purpose of these points was to provide high-contrast screen space coordinates that could be combined with their world space coordinates to somehow calibrate camera details, which we expected would become particularly complex for modeling when the camera rotated throughout a sequence. We later created a three-dimensional coordinate system and measured the world space coordinates of these yellow fiducial points using a tape measure. We recorded the center of each point, as the ?points? were actually small areas to make them detectable from the recovered film footage. Calculated camera parameters can then be used to model the entire scene as it took place while filming.

Calibration of the camera location and rotation was achieved using an algorithm known as Tsai?s method. First derived by Roger Tsai at IBM [5] and expounded upon by Lucasfilm [6] and others, this method is used frequently and open source versions are available on the Web. We used Tsai?s Camera Calibration Software Package v. 3.0, a C program maintained by Reg Wilson of CMU, to determine these values. This code is versatile and allows for simple shortcut approximations, more precise calculations, coplanar points, and noncoplanar points. The inputs to the method we used, invoked by 'nccal', should be eight or more points with complete world space (mm) and screen space (pixels)

coordinates that are noncoplanar and make sense in a right handed world-space coordinate system [7]. We performed several of these operations on simple bitmap decompressions of frames and found reasonable but less than beneficial results. Certainly the crudeness of our means of measuring the fiducial points had an impact on the results. Perhaps sampling much more points would have been helpful. As it was we had fifty-three reference points in our shooting area, but what mattered was how many were in each frame. Eight was the minimum required to solve for the unknowns of camera distortion, focal length, etc., and translation and rotation about the world space axes and origin. More would probably have lessened the standard deviation of the error terms, yielding better results. Had we the time and money we could have used more precise measuring equipment and a much more detailed analysis of the corridor to yield better results. For example, Tsai?s method calculations can be improved when one calibrates his or her own camera's filming properties, such as focal length and distortion, with very sophisticated tools. As it was, we just left these parameters unknown and let our fiducial points estimate their values.

As Poser supported modeling Fred directly into the live video sequences, we decided to use our Tsai?s method calculations in modeling only as a rough guide for placement of Fred with respect to the camera world space figures. Human perception outperformed our engineering accomplishments in this regard, as simply modeling Fred directly into the scene was rather convincing. Had we the time, we could have modeled the entire corridor and the camera's position in each scene and put Fred exactly where we wanted.

Compositing

Compositing the animated character into the live video presented another obstacle. Compositing deals with placing the completed model of Fred in the scene. It may be broken into two primary challenges with respect to this project-determining occlusion and antialiasing. The first challenge is determining occlusion of the model by objects in the live video and vice versa. Most occlusion took place between Godfrey and Fred. Another challenge arises when dealing with occlusion along the border between Fred and the live video. One way of antialiasing along these borders is to select between images in a proportional manner.

One of our original ideas was to model Godfrey into the scene with Fred and use that to determine who was in front of whom. However, this idea was discarded partially due to time constraints surrounding the complexity of creating a rough model of Godfrey. Another method of selecting between scenes for each pixel is through rgbaz proposed by Duff[3] and Porter and Duff[4]. In this method, each pixel in each of the two scenes would be given an rgbaz vector of color composition (rgb), the fraction of the picture that color covers (a), and the eye space z coordinate (z). To use this method, we would have modeled Fred with a background at some location near negative infinity. We would also have had to figure out for each pixel in the live video the right values for the alpha and z channels, effectively modeling all unoccluded objects found in the live video. Using the rgbaz vectors for each of the live video model and our Fred model, and we could combine them into an output frame buffer with the appropriate rgb values at each pixel given the alpha channel and z coordinate values from each scene. Figuring out that much detail about the live sequences and the world space coordinates of everything visible in them would prove quite a time-consuming endeavor. Even an extensive network of interpolating between points would be unwieldy and rather unhelpful given our crude measuring capabilities. As these options seemed unnecessarily time consuming, again we were able to take advantage of the limited acuteness of human perception.

Other challenges which we could have tackled in compositing but didn't include the effects of ray tracing Fred into the scene to a depth greater than one. For example, the lighting of Fred could be affected by the photons reflecting off his surroundings, and vice versa. Issues such as these were deemed to have negligible effects as the surroundings weren?t very reflective.

Shadowing

Creating proper shadowing effects in the final product proposed yet another task. An accurate model of Fred?s surroundings, stationary and mobile, was infeasible as mentioned earlier. The positioning and orientation of the lights in each scene were recorded along with the fiducial points, and from them we could estimate their relative positions with respect to the Fred?s object space from our camera position. Thus, shadowing effects of Fred on himself were handled by the modeling program.

This left shadowing of objects in the live video on Fred and shadowing of Fred on parts of the live video scene. The two lights were situated in such a way as to prevent shadows from objects other than Godfrey being cast on Fred. In addition, the shadowing of Godfrey onto Fred and of Fred onto Godfrey was taken to be out of the scope of this project, as this would involve the ordeal of modeling Godfrey, an idea which was rejected. We took these effects to be minimal or insignificant to the audience. We chose to only concern ourselves with adding shadows cast by Fred on his surroundings. Furthermore, we defined surroundings to be only objects covering large areas, modeling the floor and ceiling into the scene with Fred where appropriate. To cast these shadows, we could have written a ray tracer to determine where shadowing occurred and scaled the background of the live video by the shadow?s effects at each pixel. Instead we used Poser to model Fred?s shadows on totally white approximations to the live surroundings. We then scaled the background of the live video by multiplying the two images (almost complete combination scene and just Fred?s shadows) together.

This approach worked well, except that in some of the scenes shadows seemed to flicker for no apparent reason. The cause of this turned out to be the resolution being set too low for the refresh rate. Upon bumping the density of pixels to 1024x1024 from a previous setting of 256x256, the problem went away.

Tools Choice and the Soundtrack

The score of Man vs. Machine was composed on a MIDI Alesis QS8 synthesizer, and then synchronized to the action with help from Freestyle 2.0 sequencer.

We have considered using Credo Multimedia's Life Forms in place of Poser before committing to one or another. However, Poser was chosen after Life Forms was found to have consistently more negative reviews than the MetaCreations product (see zdnet's comparison [11] for example). Analogously, Final Cut Pro was chosen after discarding Adobe's Premier and AfterEffects alternatives.

Our key compatibility challenge was inability to work with quicktime format on MIT's Sun Workstations. This difficulty remains, as we were pressured to do all our modeling on Macs.

Individual Contributions

All modeling tasks, except for the shadowing implementation was accomplished by Josh Glazer. The composer of the soundtrack was Josh as well. All coding, compositing and Tsai's method research, write-up, and shadow modeling were done by Ken McCracken and Alex Sverdlov. Filming, storyboard development, presentation planning and preparation, including the general proof-reading and tidying up of things, were all tackled by a group effort.

Lessons Learned

We learned how to use a number of tools while spending many an hour on this project. Most notably Poser and Final Cut Pro, which were the cornerstones in our modeling efforts, as well as outdated Borland compilers (for Tsai's code), and clunky image converters among others.

Had we the freedom to attempt this project with our present experience, some approaches would have

been different. We would, for instance, take the 'mask out the ookie' route, rather than have Godfrey fight thin air, and then try to superimpose Fred. Albeit being the more laborious alternative, it would have allowed us to circumvent the problem of moves discrepancy from take to take. This was a serious challenge, since we knew how a real human body reacted to a technique, but could not follow that guideline, placing us in a position of either having to abandone an otherwise healthy scene, or manipulate it for hours trying to tweak our way to realism.

We learned that while our pick of location worked out for the reasons described earlier, a place with a different lighting might have been a wiser choice. It turned out that Poser does not support radiocity[2] modeling. However, the overhead fluorescent lights are present in many shots. If we were to do this again from start to finish, a location that did not suffer from this complication would have been used.

We learned that in sound effect dubbing, the effect is actually started a few frames before the action it is supposed to accompany. To us this delay seemed counter-intuitive, but it was clearly the way to make the effects work properly. We are still guessing as to why this would be the case.

We have learned that exhausing research of location, as well as of algorithms we were planning to use -- prior to filming -- is vital. Learning this the hard way, resulted in some frames having fewer than eight points required by Tsai's calibration code. Nonetheless, we have been able to use Tsai's technique[5], albeit not apply it to all the scenes. Learning this algorithm, as well as the compositing rgbaz method [3, 4], and applying our understanding of them to the task at hand, was both interesting and useful.

We also learned much about the innards of various image formats, and a thing or two about image compression. Interestingly, many of the formats that we explored have been designed with C++ data structures in mind for ease of use (specifically the Filmstrip[8] and FLIC[7] uncompressed file formats).

From working together as an engineering team intent on solving a single problem, we learned to break-up the challenges according to the competency areas of each team member, and to communicate effectively among ourselves, as well as with staff of 6.837, in meeting these challenges.

On a more general level, working on Man vs. Machine made us realize how truly daunting is the task of implementing accurate machine vision; as well as making us develop new appreciation for a human brain's ability to process the same information successfully. An intriguing flipside to this realization came from pondering the fact that the brain can still be tricked quite easily into misinterpreting the data, as long as a few basic principles are exploited in image construction.

Acknowledgements

Our team would like to thank Prof. Teller and Charles Lee for their feedback on our goals and progress. Their suggestions helped us understand the tasks at hand, as much as their pointers to relevant information helped us accomplish them. We especially thank Sensei Godfrey Inniss for spending his hours fighting windmills in the dustiest basement that any of us have ever seen, instead of wrestling with his own final project.

In our efforts, we have also relied heavily on work done by Roger Y. Tsai, and the detailed information on image formats compiled by CompuPhase Automatisering of Netherlands on their website [12].

Finally, we would like to thank Apple(tm) for making some of the best tools of the trade.

Bibliography

1. Hearn, Donald and Baker, M. Pauline. "Ray-Tracing Methods." *Computer Graphics: C Version.* Prentice Hall: Upper Saddle River, NJ, 1997. pp. 527-543.

2. "Radiosity." *MIT Fall 1999 6.837 Lecture Notes #17*. http://graphics.lcs.mit.edu/classes/6.837/F99/lectures/L14.ps.
3. Duff, Tom. "Compositing 3-D Rendered Images." AT&T Bell Laboratories. *Computer Graphics: 19(3)*. ACM: July 22-26, 1985. pp. 41-43.
4. Porter, Thomas and Duff, Tom. "Compositing Digital Images." Lucasfilm Ltd. *Computer Graphics: 18(3)*. ACM: July 1984. pp. 253-259.
5. Tsai, Roger Y. and Huang, Thomas S. "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses." *IEEE Journal of Robotics and Automation: RA-3(4)*. IEEE: August 1987. pp. 323-343.
6. Tsai, Roger Y. and Huang, Thomas S. "Analysis of 3-D Time Varying Scene." *IBM Research Papers*. IBM: 1982. 12pp.
7. Tsai, Roger and Wilson, Reg. "Tsai Camera Calibration Software." http://www.cs.cmu.edu/afs/cs.cmu.edu/user/rgw/www/TsaiCode.html.
8. "The FLIC File Format. Compuphase. http://www.compuphase.com/flic.htm.
9. "The Filmstip File Format. Compuphase. http://www.compuphase.com/filmstrp.htm.
10. Meta Creations Web site. Meta Creations. http://www.metacreations.com/products/poser4.
11. *MacUser*. ZDNet: May 1997. http://macuser.zdnet.com/mu_0597/reviews/review04.html.
12. CompuPhase Automatisering Web Site. CompuPhase. http://www.compuphase.com/index.html.

Appendix

A0. Screenshots:

Figure 1 -- The Cover (Fred’s introduction shot):

Figure 2 -- Mask of a pipe, used in hiding Fred’s foot underneath:

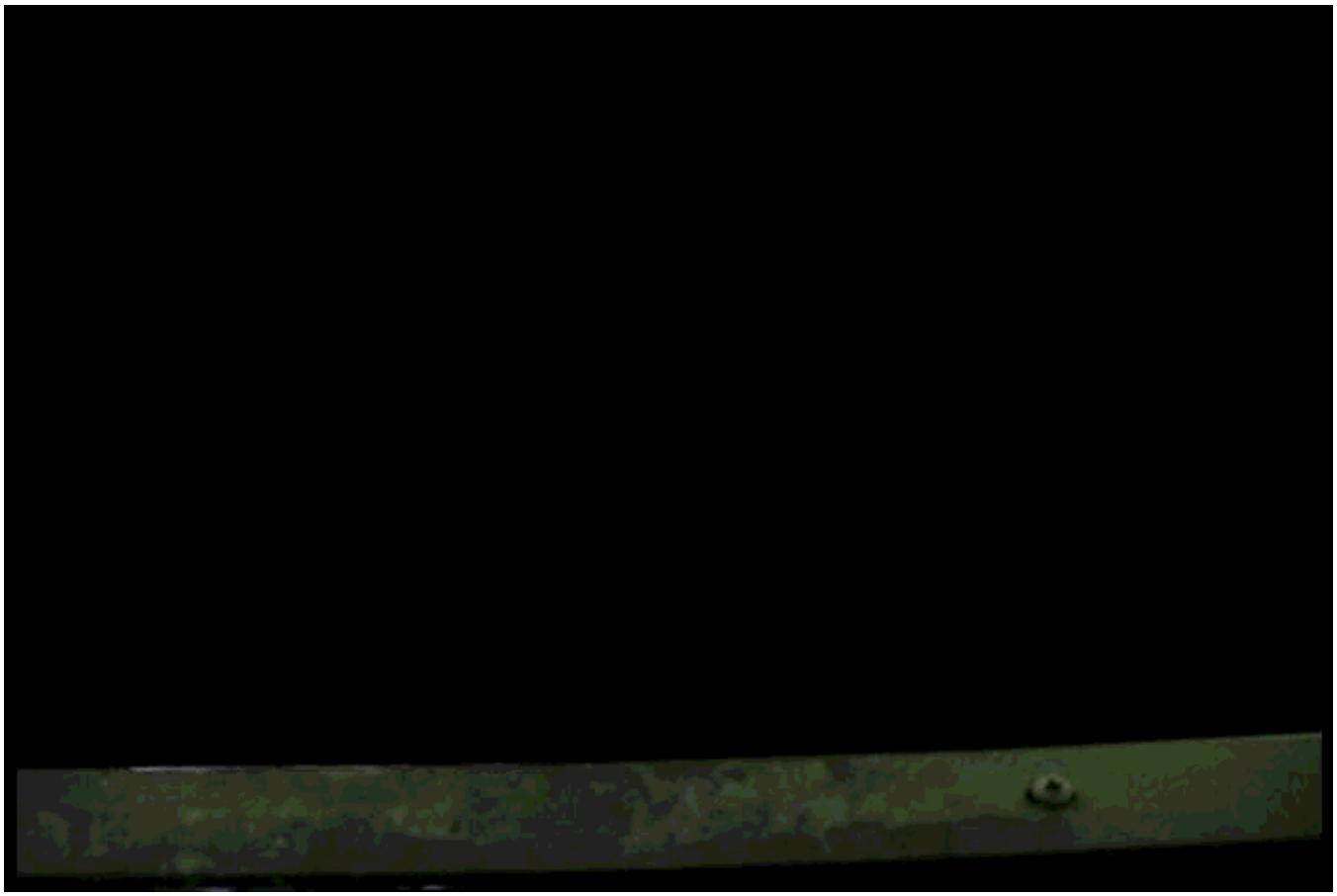

Figure 3 -- Finished version of the frame, with Fred’s foot hidden (but no shadow):

Figure 4 -- Fred’s face closeup:

Figure 5 -- Fred speaking (note lip movement):

Figure 6 -- Fred draws sword (also, eyebrow movement, blinking,and speech):

Figure 7 -- A screen capture of Poser’s environment (click to view full size):


Fred Sword 1
Right Shoulder
Options