Mapping Motion: The Principles of Motion Capture and the Law of Projection

Kelly Hamilton  
Saint Mary's College  
Notre Dame, IN 46556, USA  
hamilton@saintmarys.edu

Motion capture is, in a sense, the ultimate escape from Flatland. Computer technology for representing 3D movements on 2D surfaces by mapping data captured from moving forms to representations on the big screens of filmmaking and small screens of computer monitors has revolutionized film animation. CGI actor/creatures, such as Peter Jackson's groundbreaking direction of Gollum in The Lord of the Rings and James Cameron's Na'vi in Avatar, mapped more and more of the human actor's performance into the CG character. I will analyze the role of motion and performance capture in these two films to determine what common principles might inform their creative techniques. In “Wittgenstein and Tufte on Thinking in 3D,” I argued that the mapping relation captured by the law of projection in Wittgenstein's Tractatus Logico-Philosophicus could contribute to our understanding of Edward Tufte’s theories of envisioning information. There, the projective relation mapping information to visual data displays is central to representation. In this talk, I will explore whether such a mapping relation, and the law of projection informing it, can also ground principles of representation in motion capture as it functions creatively in filmmaking.


1. INTRODUCTION

How do you project creative performances involving motion in three spatial dimensions into a flat surface without losing essential elements of the artistic performance? Computer graphics opened new aesthetic possibilities, transforming major art forms. Motion capture has a long history, but when it was translated into computer graphics in the late twentieth century, the effect was transformative.

Motion capture is an important technique with many uses, but its potential for creative filmmaking has only recently been realized in ground-breaking movies.

Motion capture is the process of recording a live motion event and translating it into usable mathematics by tracking a number of key points in space over time and combining them to obtain a single three-dimensional (3D) representation of the performance. The captured subject could be anything that exists in the real world and has motion (Menache 2011, 2).

In motion capture (mocap), markers at many points on a moving body provide data about the body's motion, which is projected into a 3D representation.

Willing suspension of disbelief is critical to an audience’s film experience, and Computer Generated Imaging (CGI) revolutionized the movie-going experience. From Gollum in The Lord of the Rings to the Na'vi in James Cameron's Avatar, motion capture and performance capture deepened artistic performances, helping create two of the most successful movie franchises of all time. In this paper, I present a theoretical framework for these artistic achievements, arguing that the method of projection in Wittgenstein's Tractatus provides us with a deeper understanding of their visual power.

2. WITTGENSTEIN, MOTION CAPTURE, AND THE METHOD OF PROJECTION

Ludwig Wittgenstein presented his famous picture theory of language in the Tractatus, arguing that in using language “we picture facts to ourselves”(2.1). For Wittgenstein, facts are “what is the case,” how things stand in relation to one another in the world. His example of a model used in a Paris courtroom to represent the situation, or what was actually the case in the matter before the court, captures this idea.

Translation of the German word Bild as picture can be misleading. (Wir machen uns Bilder der Tatsache), David Stern warns

Wittgenstein used the German word 'Bild' to talk about the model, a term usually translated as
‘picture’; as a result, the theory of meaning it inspired is generally known as the picture theory. While both words cover such things as images, film frames, drawings, and paintings, the idea of a three-dimensional model is more readily conveyed by the German ‘Bild’ than the English ‘picture’. It is important not to be misled: the theory involves generalizing from what models, pictures, and the like are supposed to have in common, and treats two-dimensional pictures as just one kind of ‘Bild’ (Stern 1995, 35-36).

In “Wittgenstein and the Mind’s Eye,” I argued that Wittgenstein’s training as an engineer at the Technische Hochschule in Charlottenberg-Berlin, then the MIT of Europe, profoundly affected his philosophy in the Tractatus. His engineering training focused on hands-on experimentation, mathematical drawing such as descriptive geometry, and engineering design. The visual thinking of the design engineer would have become second nature.

Wittgenstein was concerned with “the logic of our language,” but he did not only mean our “two dimensional script.” He was concerned with the languages of all of our modes of representation. He uses music as the illustration of his law of projection.

4.014 A gramophone record, the musical idea, the written notes, and the sound waves, all stand to one another in the same internal relation of depicting that holds between language and the world. They are all constructed according to a common logical pattern.

4.0141 There is a general rule by means of which the musician can obtain the symphony from the score, and which makes it possible to derive the symphony from the groove on the gramophone record, and, using the first rule, to derive the score again. That is what constitutes the inner similarity between these things which seem to be constructed in such entirely different ways. And that rule is the law of projection which projects the symphony into the language of musical notation. It is the rule for translating this language into the language of gramophone records.

Projection enables us to translate from the musical idea, to the written notes, to the groove on the gramophone record, and it does that because what is projected is the logical form, the internal pattern of depicting. “The possibility of all imagery, of all our pictorial modes of projection, is contained in the logic of depiction” (4.015).

What is important for our argument in this passage is the law of projection and its relation to the logic of depiction. Some knowledge of projective geometry is helpful to understand how its manner of representing 3D bodies in space is reflected in the doctrines of the Tractatus. Technical terms from projective geometry appear in the Tractatus, and these ideas are also an excellent description of techniques involved in motion capture.

Projection has been described as a valuable method for investigating the mathematical properties of figures. Descriptive geometry is “the geometry in which figures are represented and their properties are investigated by means of projection” (Willson 1909, 7). Assuming a center of projection, S, and a point in space, A, the straight line connecting S with A is a projecting line (also called a ray or projector). “The word projection is used not only to indicate the method of projection but also the representation itself” (Willson 1909, 2).

Thus, points on the surfaces of figures, represented in a mathematical space where relative position is determined with absolute accuracy, are connected to real world figures. When the subject of these mathematical drawings is an invention that fails, or a working bridge, their realism is obvious. Wittgenstein’s graduate research in aeronautical engineering led to a patent for a propeller. Engineers think, design, and invent using mathematical drawings.

In the Tractatus, Wittgenstein’s Bilder stand in a “projective relation” to reality. “A Bild can depict any reality whose form it has” (2.171). For “a Bild is a model of reality” (2.12). In its projective relation to the world, “a proposition is a model of reality as we imagine it” (4.01). “We use the perceptible sign of a proposition (spoken or written, etc.) as a projection of a possible situation” (3.11). “A proposition is a propositional sign in its projective relation to the world” (3.12).

Wittgenstein’s law of projection involves the visual thinking of the engineer and the descriptive geometer. If you think of the sense of the proposition, you will see it in its projective relation to reality. The visual thinking of the engineer involves seeing projective relations in the mind’s eye.

Seeing the Bild theory from this perspective is unusually concrete, but makes sense. For, in Wittgenstein’s words, “Our problems are not abstract, but perhaps the most concrete that there are” (5.5563). So, how does the method of projection involved in the visual thinking of the engineer help us understand the role of motion capture in modern filmmaking?
3. THE MAKING OF GOLLUM

Filming Tolkien’s *The Lord of the Rings* was going
to be difficult for many reasons. One of the most widely read books of the twentieth century, the classic trilogy had a devoted following. Tolkien, a philology professor at Oxford, crafted a complex world with a deeply layered history and sophisticated cultures, posing a formidable challenge to film representation.

Many creatures of Middle Earth would be difficult to bring to the screen, and one of them was central to the dramatic story. Jackson knew Gollum was crucial. “If Gollum wasn’t believable…you were going to kill not just *The Two Towers*, but *The Return of the King*” (Jackson 2001). Getting Gollum right was essential.

Gollum was originally conceived as a CG character, and the art used to create the rig for his character came first. The rig is critical in the creation of CG characters. As explained in *Thinking Animation: Bridging the Gap between 2D and CG*, which describes the transition from traditional key frame animation to computer-generated animation, the rig is vital to the creation of a CG character. It is through the rig that the animated character acts, performing its role in the story.

The one important tool you need to create good animation is your rig. The rig is the puppet you use to animate with on the computer…Your rig is your character…Working with your rigger, you can create the most important tool a CG animator can have—a solid puppet. With a clean and simple model design and a solid rig, you are on your way to creating a memorable character (Jones and Oliff 2007).

Most of the conceptual design for Gollum’s rig came from the sculpting room. Weta Workshop sculpted a scannable maquette for Weta Digital’s rig for the CG Gollum.

Enter Andy Serkis, who auditioned for the voice of the CG Gollum. His performance changed Jackson’s conception of Gollum’s role, as his acting on the set would change his direction of Gollum’s performance.

In order to create the voice, he was having to …put all this expression on his face and that was where he was finding the voice. He was actually doing the character…In that audition, I came to realize something that had never really occurred to me—the voice, and the facial expressions…are related (Jackson 2001)

Serkis, a Shakespearean actor, was hired as the voice actor. He actually became the personality of Gollum, “the motivating, unifying performance of his character” (Jackson 2001).

Serkis performed with the other actors as an animation reference, wearing a lycra suit in which his performances were shot. Then the scenes were shot without him, a mime pass in which they pretended he was there for placement of the CG character. Initially, the reference pass was not taken seriously, as the mime pass would be the movie performance.

The performances, however, were better when Serkis was in the scenes. As Jackson said, “we gradually learned we really couldn’t do without him.” For one thing, “the acting was always better on the Andy plates, you didn’t get that on the mime pass” (Jackson 2001). Serkis began to drive Gollum’s character with the expressiveness of his performances.

Jackson had the CG Gollum redesigned, scanning Serkis’ face into the computer and building a new sculptural maquette based on his features to represent the actor’s expressions on the rig. It meant redoing animation in production, which was expensive, difficult, and truly worth the effort. Even so, the actor’s face would still be created with key frame animation.

Jackson wanted Andy’s performance to move the puppet, so they used motion capture to film the performance. Motion capture involved placing dots on a body suit worn by the actor as reference points. They would be picked up by cameras all around the set, recording the reflected light to create data for reconstructing his movements on the computer screen.

It happened in real time, allowing Jackson to direct Serkis performing as Gollum through a 3D representation of his performance on a screen. Seeing a rough model of the performance in real time on the display was an advantage over traditional animation. As Jackson put it, “you could look over at Andy doing the scene, then you could look down at the Gollum character on the screen doing the same thing” (Jackson 2001). They could shoot different takes, and the character was rendered overnight for dailies the next morning.

Gollum’s animation generated tension between the traditional animators and the digital artists working with motion capture data, however, reflecting tensions in the animation community. Traditional key frame (2D) animation and computer-generated (CG or 3D) animation involved different mediums with distinct strengths and weaknesses. Both types of animation were used to create Gollum.

The computer had changed an art form that had been a pencil and paper medium for most of the twentieth century.
Animators find themselves in the midst of a momentous change in their industry. As in many other fields, the computer has made what is known as a disruptive impact on our art form. Think of the car and the horse, the cellular phone and the pay phone, the CG and 2D feature (Jones and Olliff 2007).

That story is summarized starkly by the authors of Thinking Animation, traditional animators writing about the transition to the “brave new world” of CG animation.

DreamWorks and Disney, the two largest animation studios in Hollywood, have shut down their traditional animation units for theatrical release, no longer accept traditional portfolios, and have dedicated all of their efforts to making CG films. In the 10-year period between the unbelievable success of The Lion King and the release of Shrek 2, 2D production has literally dwindled to nothing. A more profound illustration of the impact of the computer on our industry would be hard to find (Jones and Olliff 2007).

The success of visual effects-driven movies in the 1990s changed the field as audiences responded to the new aesthetic. “The sheer richness of 3D and its ability to move the camera around in this new world made traditional animation seem, quite literally, flat” (Jones and Olliff 2007). Profits for traditionally animated films fell, and CG films were tremendously successful.

By 2001, the 2D boom was in the past, and the success of Shrek was central to changing the success of traditional animation. Shrek, Toy Story, and other CG films proved that grounding yourself equally in the arts and in computer sciences was the key to staying employed (Jones and Olliff 2007).

These two sets of accomplished artists with distinctive skills, techniques, and artistic sensibilities were only beginning to understand how they could work together.

Drawing skills are important for animation.

(This) is a heavily debated topic among stop-motion, 2D, and CG animators alike. The consensus is that everyone needs to be able to draw well. Even if it’s a stick figure, drawing what you envision for the scene helps you understand things...Drawing helps you think more logically and clearly about where the scene should go. Drawing helps you see the whole picture...Drawing is a form of communication that rises above what words alone can accomplish, and the old adage that a picture is worth a thousand words is truer than ever (Jones and Olliff 2007).

In “Wittgenstein and the Mind’s Eye,” I argue that drawing is crucial as a language for visualization. Drawings embody an engineer’s knowledge.

Engineering design draws on that skill and intuition. “Engineering drawings are expressed in a graphic language, the grammar and syntax of which are learned through use; it also has idioms that only initiates will recognize” (Ferguson 1992, 3).

In A History of Engineering Drawing, Peter Booker expands on this thought:

In its narrowest sense engineering drawing is a language used for communication. However, languages in general are not only useful for communication; they play an inherent part in our very thinking, for we tend to think in terms of the languages that we know. Drawing is of this nature, and he who can draw can think of, and deal with, many things and problems which another man can not. Between thinking and communication, in the form of geometry, drawing has another function; . . . it is . . . a primary tool of design (Booker 1963, xv).

His theme is “the representation of three-dimensional objects on a two-dimensional surface” (Booker 1963, xv). Drawing skill, acting with a pencil, was as important as motion capture for creating Gollum.

These animating communities initially clashed when creating Gollum. Ultimately, they blended motion capture performance with key frame animation. They didn’t use motion capture for facial work, for example. Traditional animators drew the facial acting. Gollum’s performance, based on the collaboration of these animators, was a “great culmination of motion capture and key frame animation” (Jackson 2001).

There was also an older technology for motion capture in animation used by Walt Disney. Rotoscope involves live action filming overlaid on a transparent frame to be traced for animation, tracking the actor’s motion frame by frame. Weta combined rotoscope and animation, referred to as rotomation. They superimposed the CG figure on the live performance. With live performances as the reference, they put CG Gollum in Andy’s place, mapping him onto the performances, then painted Andy out.

Serkis captured Gollum’s twisted pathos, at once evil, manipulative, and pathetic, and motion capture of his athleticism and physical expressiveness gave believability and depth to the character. He discussed Gollum’s character with the design team, changing their understanding of his character. Smeagol, after all, had become Gollum. Sinister as he was, he was not a one-dimensional villain. As Gandalf said. Smeagol’s was a sad life. Once he saw him, Frodo did pity him. Serkis’ performance captured what the ring had done to Smeagol, and his struggle with the remnants of his better self.

234
The last piece in the CG character’s transformation was the decisive scene where Smeagol struggles with the voice in his head he cannot drown out, and triumphs. The schizophrenia scene was masterfully acted by Serkis as a one shot. His portrayal of competing personalities in that intense exchange was flawless, one of the finest pieces of acting in the film. Smeagol was free of Gollum. The team commented that there were some scenes where they used no Andy reference at all, and some scenes where they used Andy references a bit, “but that was one where we used it to the letter” (Jackson 2001).

The performance was projected onto the Gollum puppet. The point to point reference of the markers on the mocap suit to the CG puppet were projective rays connecting Andy’s performance to Gollum’s movements onscreen. The dots sending information about movement in three-dimensional space instantiated Wittgenstein’s law of projection, illustrating the method of depiction and showing how the inner similarity between the performances was preserved. The logic of depiction enabled the patterns of depicting to translate from the live performance to the CG character. CG Gollum, informed by the performance of Andy Serkis, was a dramatic success. Jackson told Serkis, “you are making film history here” (Jackson 2001). The performance was a milestone in film animation. Then James Cameron took performance capture to a whole new level in Avatar.

4. JAMES CAMERON’S NA’VI: “BUILDING THE PARACHUTE ON THE WAY DOWN”

James Cameron is larger than life in more ways than one, combining unusual talents with a drive to continually test the limits of the possible. Reflecting the talents and interests of his mother the artist and his father the engineer, Cameron loved to draw, tinker, and build from childhood. These traits would shape his outstanding success as a filmmaker.

Although work on Avatar did not start until 2005, Cameron had written a scriptment for it in 1995. What he wanted to do was impossible then, so he set it aside and waited for the technology to mature. In 2005, CG human facial performance still wasn’t emotionally real enough for him. There had been progress, however, and promising performances such as Gollum.

In 2005, Cameron began a year of research for Avatar. He was pushing the envelope — again. There were two sides of the research project: a technical and design side. Designing the strikingly beautiful flora and fauna of his alien world was difficult; it took almost two years to get a finished banshee. He kept asking his artists, what is the metaphor? What are we communicating with every bone and sinew that we put into this creature?

The design of Pandora began with flat artwork. With nothing tactile for reference, they turned to their sculptors. Sculpting maquettes provides three-dimensional, tactile art, a basis for developing character and capturing the personalities of the animated characters in their rig/puppets. Weta Digital created the computer animation. As with Gollum, the actors’ facial features were used for the CG faces.

Cameron also assembled his technical system. In Brother Termite, he had already used facial performance capture as an image based process. By 2007 he was ready. The screen performances would be done by the actors, and designing their rigs was both crucial and difficult.

Most of James Cameron’s space epic, Avatar, was shot on a performance-capture stage, known as the volume, in Playa Vista, Calif. The volume was rimmed by 120 stationary video cameras, which could record the movements of all actors at once in 3D, with submillimeter precision. Data from the cameras was streamed into Autodesk software, which translates actors’ movements into digital characters in real time within a low-resolution computer-generated environment. So riding a fake banshee mock-up onstage instantly translated to CG footage. Multiple cameramen were used on set for reference video, but because the volume essentially captures performances from every angle at once, Cameron could digitally render whatever angles and shots he wanted after the performance, adjusting the camera movements while viewing playback (Thompson 2009).

What Cameron saw as he directed in real time in the volume was very rough animation. In a spatial volume with markers on the actors’ body picked up by 100s of cameras, however, you can construct a CG character in real time. You can direct CG acting.

Training to be Na’vi was intense. Actors learned to move like the Na’vi with a coach. When Cameron told people that Zoe Saldana had to train for months to be Na’vi, they laughed. Why was that necessary for an animated film? His response was that it wasn’t animated; it was captured performance. The physicality of her character, and this applied to all the Na’vi and avatars, was created by the actor. Performances had the culturally specific grace of Na’vi movements and translated faithfully to the screen through performance capture. What happened in the volume went on the screen.
In their sandbox spatial environment, they were creating digital characters structured by data from the actors’ performances. Performance capture recorded movement, which was translated into the digital framework in real time. Cameron saw the performances. He had the freedom to see his takes and direct his actors.

Programming facial muscles to work was different for each character. They scanned multiple copies of the faces into the computer. For facial performances, they mounted a camera directly to the actors’ head. Instead of makeup, the actors’ faces were painted with multiple small dots that the head rig camera used to record every movement. The head rig was a kind of skull cap, based on a life cast and laser scan of the actors’ head, that fit tightly, smoothly, and comfortably. A carbon boom held a little camera to shoot the face, capturing the facial performance, recording every facial movement. That level of detail proved to be the “Holy Grail” approach to creating CG faces. Cameron said

“I knew I could not fail if I had a 100 percent closeup of the actor 100 percent of the time that traveled with them wherever they went’...The information from the cameras produced a digital framework, or rig, of an actor’s face. The rig was then given a set of rules that applied the muscle movements of each actor’s face to that of the avatar or the Na’vi that he or she was playing. To make a CG character express the same emotion as a human actor, the rig had to translate every arch of a human eyebrow directly to the digital character’s face (Thompson 2010).

The first efforts at performance capture were still inadequate.

After Cameron complained about the uncanny-valley effect, Weta spent another year perfecting the rig on Worthington’s avatar by tweaking the algorithms that guided its movements and expressions until he came alive enough...when Weta was finished, you could pour the motion-capture data into the rig and it would come out the other side right (Thompson 2010)

Cameron restricted the animation artists to drawing with performance capture data. They were only allowed to add movements for the tails and the ears, which the actors could not perform. Neytiri’s ears flattening and tail twitching are examples of the final animation of the Na’vi, who were not, after all, human. Cameron described the specifically Na’vi character elements as a turbo charger on top of the performance.

What was viewed on the screen was the undiminished human performance. The digital world took on the character of the real world. The animators did not create the performances; the actors did. The actors in the volume knew that what mattered was their performances. What they did in the volume was the definitive human emotion on the movie screen.

Dots on actors’ faces and markers on their bodies were used to record their movements, capturing detailed data about their performances. That information informed CG characters, creating a moving, acting 3D model performing in virtual space. The metaphor for that process is the projective ray of mathematical drawing. Information from those dots and markers mapped human performances onto CG characters on the movie screen, representing their performances with amazing integrity.

In Wittgenstein’s words, the inner similarity of the performances was preserved. Developing the technology to make this level of performance capture possible, Cameron could tell the story of Avatar. It was a different level of challenge and accomplishment, even for him.

5. MAPPING MOTION: PERFORMANCE CAPTURE AND THE LAW OF PROJECTION

How do we represent our three-dimensional world in flatland? Using motion and performance capture to project live performances onto the movie screen, Cameron pushed artistic innovation in film to its limits. He produced a film that is an enduring work of art at the same time it is a record breaking box office success.

The underlying metaphor for this process is projection. As Elizabeth Anscombe said,

Wittgenstein’s use of “projection” is a metaphorical extension of the mathematical use, which may be explained thus: “The drawing of straight lines through every point of a given figure, so as to produce a new figure each point of which corresponds to a point of the original figure.” The new figure is also said to be a projection of the original one, which is projected into it (Anscombe 1959, 69).

That powerful insight captures the core vision of the Tractatus and illuminates the artistic process of creating characters such as Gollum and movies such as Avatar.

Motion capture and performance capture preserve the inner similarity between live performances and their CG realization. Point to point correspondences between markers and dots on the actors’ bodies and the CG characters translate performances faithfully. Cameron insisted that “these are actor’s performances. It is not a visual
effect. These people have played these characters."

The combination of engineer and artist in James Cameron would have appealed to Wittgenstein, an engineer and an architect who came from a family of legendary art patrons. His philosophical foundations were grounded in the artistic world of fin-de-siècle Vienna and the Krausian critique of all our modes of representation (Janik Toulmin 1973). He would have understood a demanding insistence on the integrity of the projection of the artistic performances to the movie screen, and he would have approved.

6. REFERENCES


