Animation

http://anm.sagepub.com

Image Future

Lev Manovich

Animation 2006; 1; 25
DOI: 10.1177/1746847706065839

Published by:
SAGE Publications
http://www.sagepublications.com

animation: an interdisciplinary journal

article

Image Future

Lev Manovich

Abstract Today the techniques of traditional animation, cinematography, and computer graphics are often used in combination to create new hybrid moving image forms. This article discusses this process using the example of a particularly intricate hybrid – the Universal Capture method used in the second and third films of *The Matrix* trilogy. Rather than expecting that any of the present 'pure' forms will dominate the future of visual and moving image cultures, it is suggested that the future belongs to such hybrids.

Keywords animation, cinematography, computer animation, computer graphics, motion capture, motion graphics, virtual cinematography

For the larger part of the 20th century, different areas of commercial moving image culture maintained their distinct production methods and distinct aesthetics. Films and cartoons were produced completely differently and it was easy to tell their visual languages apart. Today the situation is different. Computerization of all areas of moving image production created a common pool of techniques, which can be used regardless of whether one is creating motion graphics for television, a narrative feature, an animated feature, or a music video. The ability to composite many layers of imagery with varied transparency, to place still and moving elements within a shared 3D virtual space and then move a virtual camera through this space, to apply simulated motion blur and depth of field effect, to change over time any visual parameter

animation: an interdisciplinary journal (http://anm.sagepub.com)
Copyright © 2006 SAGE Publications (London, Thousand Oaks, CA and New Delhi)
Vol 1(1): 25–44 [1746-8477(200607)]10.1177/1746847706065839

of a frame – all these can now be equally applied to any images, regardless of whether they were captured via a lens-based recording, drawn by hand, created with 3D software, etc.

The existence of this common vocabulary of computer-based techniques does not mean that all films now look the same. What it means, however, is that while most live action films and animated features do look quite distinct today, this is the result of deliberate choices rather than the inevitable consequence of differences in production methods and technology. At the same time, outside the realm of live action films and animation features, the aesthetics of moving image culture dramatically changed during the 1990s.

What happened can be summarized in the following way. Around the mid-1990s, the simulated physical media for moving and still image production (cinematography, animation, graphic design, typography), new computer media (3D animation), and new computer techniques (compositing, multiple levels of transparency) started to interact within a single computing environment – either a personal computer or a relatively inexpensive graphics workstation affordable for small companies and even individuals. The result was the emergence of a new hybrid aesthetics that quickly became the norm. Today this aesthetics is at work in practically all short moving image forms: TV advertising and TV graphics, music videos, short animations, broadcast graphics, film titles, music videos, web splash pages. It also defines a new field of media production – motion graphics – but it is important to note that the hybrid aesthetics is not confined to this field but can be found at work everywhere else.

This aesthetics exists in endless variations but its logic is the same: juxtaposition of previously distinct visual languages of different media within the same sequence and, quite often, within the same frame. Hand-drawn elements, photographic cutouts, video, type, 3D elements are not simply placed next to each other but interwoven. The resulting visual language is a hybrid. It can also be called a metalanguage as it combines the languages of design, typography, cell animation, 3D computer animation, painting, and cinematography.

In addition to special effects features, the hybrid (or meta) aesthetics of a great majority of short moving images sequences that surround us today is the most visible effect of computerization of moving image production. In this case, animation frequently appears as one element of a sequence or even a single frame. But this is just one, more obvious, role of animation in the contemporary post-digital visual landscape. In this article I will discuss its other role: as a generalized technique that can be applied to any images, including film and video. Here, animation functions not as a medium but as a set of general-purpose techniques – used together with other techniques in the common pool of options available to a filmmaker/designer.

I have chosen a particular example for my discussion that I think illustrates well this new role of animation. It is a relatively new method of combining live action and CG. Called 'Universal Capture' (U-cap) by their creators, it was first systematically used on a large scale by ESC Entertainment in the *Matrix 2* (2003) and *Matrix 3* (2003) films from *The Matrix* trilogy. I will discuss how this method is different from the now standard and older techniques of integrating live action and computer graphics elements. Universal Capture also creates visual hybrids – but they are quite different from the hybrids found in motion graphics and other short moving image forms today. In the case of Universal Capture, different types of imagery are not mixed together but rather *fused* to create a new kind of image. This image combines 'the best' qualities of two types of imagery that we normally understand as being ontological opposites: live action recording and 3D computer animation. I will suggest that such image hybrids are likely to play a large role in future visual culture while the place of 'pure' images that are not fused or mixed with anything is likely to diminish.

Uneven development

What kinds of images are likely to dominate visual culture a number of decades from now? Will they still be similar to the typical image that surrounds us today – photographs that are digitally manipulated and often combined with various graphical elements and type? Or will future images be completely different? Will photographic code fade away in favor of something else?

There are good reasons to assume that future images are likely to be photograph-like. Like a virus, a photograph turned out to be an incredibly resilient representational code: it survived waves of technological change, including computerization of all stages of cultural production and distribution. The reason for this persistence of the photographic code lies in its flexibility: photographs can be easily mixed with all other visual forms – drawings, 2D and 3D designs, line diagrams, and type. As a result, while photographs truly dominate contemporary visual culture, most of them are not pure photographs but various mutations and hybrids: photographs that went through various filters and manual adjustments to achieve a more stylized look, a more flat graphic look, more saturated color, etc.; photographs mixed with design and type elements; photographs that are not limited to the part of the spectrum visible to a human eye (night vision, x-ray); simulated photographs done with 3D computer graphics; and so on. Therefore, while we can say that today we live in a 'photographic culture', we also need to start reading the word 'photographic' in a new way. 'Photographic' today is really photo-GRAPHIC, the photo providing only an initial layer for the overall graphical mix. (In the area of moving images, the term 'motion graphics' captures perfectly the same development: the subordination of live action cinematography to the graphic code.)

One way in which change happens in nature, society, and culture is inside out. The internal structure changes first, and this change affects the visible skin only later. For instance, according to the Marxist theory of historical development, infrastructure (i.e. mode of production in a given society – also called 'base') changes well before superstructure (ideology and culture in this society). In a different example, think of technology design in the 20th century: typically a new type of machine was at first fitted within old, familiar skin: for instance, early 20th-century cars emulated the form of the horse carriage. The familiar McLuhan idea that new media first emulate old media is another example of this type of change. In this case, a new mode of media production, so to speak, is first used to support the old structure of media organization, before the new structure emerges. For instance, the first typeset books were designed to emulate handwritten books; cinema first emulated theatre; and so on.

This concept of uneven development can be useful in thinking about changes in contemporary visual culture. Since this process started 50 years ago, computerization of photography (and cinematography) has by now completely changed the internal structure of a photographic image. Yet its 'skin', i.e. the way a typical photograph looks, largely remains the same. It is therefore possible that at some point in the future the 'skin' of an image will also become completely different, but this has not happened yet. So we can say at present our visual culture is characterized by a new computer 'base' and an old photographic 'superstructure'.

The trilogy of *Matrix* films provides us with a very rich set of examples that are perfect for thinking further about these issues; it is an allegory about how its visual universe is constructed. That is, the films tell us about *The Matrix*, the virtual universe that is maintained by computers – and, of course, visually the images of *The Matrix* which we the viewers see in the films were all indeed assembled with the help of software (the animators sometimes used Maya but mostly relied on custom-written programs). So there is a perfect symmetry between us, the viewers of a film, and the people who live inside *The Matrix* – except that, while the computers running *The Matrix* are capable of doing it in real time, most scenes in each of *The Matrix* films took months and even years to put together. (So *The Matrix* can be also interpreted as a futuristic vision of computer games at a point in the future when it becomes possible to render *The Matrix*-style visual effects in real time.)

The key to the visual universe of *The Matrix* is the new set of computer graphic techniques that over the years were developed by a number of people both in academia and in the special effects industry, including Georgi Borshukov and John Gaeta.[1] Their inventors coined a number of names for these techniques: 'virtual cinema', 'virtual human', 'virtual cinematography', 'universal capture'. Together, these techniques represent a true milestone in the history of computer-driven special effects. They take to their logical conclusion the developments of the 1990s, such as motion capture, and simultaneously open a new stage. We can say that with *The Matrix* (1999), the old 'base' of photography has finally been completely replaced by a new computer-driven one. What remains to be seen is how the 'superstructure' of a photographic image – what it represents and how – will change to accommodate this 'base'.

Reality simulation versus reality sampling

Before proceeding, I should note that not all of the special effects in *The Matrix* rely on Universal Capture and, of course, other Hollywood films already use some of the same strategies. However, in this article I focus on the use of this process in *The Matrix* because Universal Capture was actually developed for the second and third films of the trilogy. And while the complete credits for everybody involved in developing the process would run for a number of lines, in this text I will identify it with Gaeta. The reason is not because, as a senior special effects supervisor for *The Matrix Reloaded* (2003) and *The Matrix Revolutions* (2003), he got most publicity. More importantly, in contrast to many others in the special effects industry, Gaeta has extensively reflected on the techniques he and his colleagues have developed, presenting it as a new paradigm for cinema and entertainment, and coining useful terms and concepts for understanding it.

In order to understand better the significance of Gaeta's method, let us briefly run through the history of 3D photo-realistic image synthesis and its use in the film industry. In 1963, Lawrence G. Roberts (a graduate student at MIT) became one of the key people behind the development of Arpanet, and published a description of a computer algorithm to construct images in linear perspective. These images represented the objects' edges as lines; in contemporary language of computer graphics they can be called 'wire frames'. Approximately 10 years later, computer scientists designed algorithms that allowed for the creation of shaded images (so-called Gouraud shading and Phong shading, named after the computer scientists who created the corresponding algorithms). From the middle of the 1970s to the end of the 1980s, the field of 3D computer graphics went through rapid development. Every year new fundamental techniques were created: transparency, shadows, image mapping, bump texturing, particle system, compositing, ray tracing, radiosity, and so on.[2] By the end of this creative and fruitful period in the history of the field, it was possible to use a combination of these techniques to synthesize images of almost every subject that were often not easily distinguishable from traditional cinematography.

All this research was based on one fundamental assumption: in order to re-create an image of reality identical to the one captured by a film camera, we need to systematically simulate the actual physics involved in construction of this image. This means simulating the complex interactions between light sources, the properties of different materials (cloth, metal, glass, etc.), and the properties of physical film cameras, including all their limitations such as depth of field and motion blur. Since it was obvious to computer scientists that if they exactly simulate all these physics, a computer would take forever to calculate even a single image, they put their energy into inventing various short cuts that would create sufficiently realistic images while involving fewer calculation steps. So in fact each of the techniques for image synthesis I mentioned in the previous paragraph are one such 'hack' – a particular approximation of a particular subset of all possible interactions between light sources, materials, and cameras. This assumption also means that you are re-creating reality step-by-step from scratch. Every time you want to make a still image or an animation of some object or a scene, the story of creation from the Bible is being replayed.

(I imagine God creating the universe by going through the numerous menus of a professional 3D modeling, animation, and rendering program such as Maya. First he has to make all the geometry: manipulating splines, extruding contours, adding bevels . . . next for every object and creature he has to choose the material properties: specular color, transparency level, image, bump and reflexion maps, and so on. He finishes one set of parameters, wipes his forehead, and starts working on the next set. Now on to defining the lights: again, dozens of menu options need to be selected. He renders the scene, looks at the result, and admires his creation. But he is far from being done: the universe he has in mind is not a still image but an animation, which means that the water has to flow, the grass and leaves have to move under the blow of the wind, and all the creatures also have to move. He sighs and opens another set of menus where he has to define the parameters of algorithms that simulate the physics of motion. And on, and on, and on. Finally the world itself is finished and it looks good; but now God wants to create Man so he can admire his creation. God sighs again, and takes from the shelf a particular Maya manual from the complete set which occupies the whole shelf . . .)

Of course we are in a somewhat better position than God was. He was creating everything for the first time, so he could not borrow things from anywhere. Therefore everything had to be built and defined from scratch. But we are not creating a new universe but instead visually simulating a universe that already exists, i.e. physical reality. Therefore computer scientists working on 3D computer graphics techniques realized early on that, in addition to approximating the physics involved, they can also sometimes take another shortcut. Instead of defining something from scratch through the algorithms, they can simply *sample* it from existing reality and incorporate these samples in the construction process.

The examples of the application of this idea are the techniques of texture mapping and bump mapping which were introduced in the second part of the 1970s. With texture mapping, any 2D digital image – which can be a close-up of some texture such as wood grain or bricks, but which can also be anything else, for instance a logo, a photograph of a face or of clouds – is mathematically wrapped around a 3D model. This is a very effective way to add the visual richness of a real world to a virtual scene. Bump texturing works similarly, but in this case the 2D image is used as a way to quickly add complexity to the geometry itself. For instance, instead of having to manually model all the little cracks and indentations that make up the 3D texture of a concrete wall, an artist can simply take a photograph of an existing wall, convert it into a grayscale image, and then feed this image into the rendering algorithm. The algorithm treats the grayscale image as a depth map, i.e. the value of every pixel is interpreted as the relative height of the surface. So in this example, light pixels become points on the wall that are a little in front while dark pixels become points that are a little behind. The result is an enormous saving in the amount of time necessary to recreate a particular but very important aspect of our physical reality: a slight and usually regular 3D texture found in most natural and many human-made surfaces, from the bark of a tree to a woven cloth.

Other 3D computer graphics techniques based on the idea of sampling existing reality include reflection mapping and 3D digitizing. Despite the fact that all these techniques were widely used as soon as they were invented, many people in the computer graphics field (as far as I can see) always felt that they were cheating. Why? I think this was because the overall conceptual paradigm for creating photorealistic computer graphics was to simulate everything from scratch through algorithms. So if you had to use the techniques based on directly sampling reality, you somehow felt that this was just temporary – because the appropriate algorithms were not yet developed or because the machines were too slow. You also had this feeling because once you started to manually sample reality and then tried to include these samples in your perfect algorithmically defined image, things would rarely fit exactly right, and painstaking manual adjustments were required. For instance, texture mapping would work perfectly if applied to a straight surface, but if the surface were curved, inevitable distortion would occur.

Throughout the 1970s and 1980s, the ‘reality simulation’ paradigm and ‘reality sampling’ paradigms co-existed side-by-side. More precisely, as I suggested earlier, a sampling paradigm was ‘embedded’ within a reality simulation paradigm. It was common sense that the right way to create photorealistic images of reality is by simulating its physics as precisely as one could. Sampling existing reality now and then, and then adding these samples to a virtual scene was a trick, a shortcut within an overwise honest game of simulation.

Building *The Matrix*

So far we have looked at the paradigms of the 3D computer graphics field without considering the uses of the simulated images. So what happens if you want to incorporate photorealistic images into a film? This introduces a new constraint. Not only every simulated image has to be consistent internally, with the cast shadows corresponding to the light sources, and so on, but now it also has to be consistent with the cinematography of a film. The simulated universe and live action universe have to match perfectly. (I am talking here about the 'normal' use of computer graphics in narrative films and not the hybrid aesthetics of TV graphics, music videos, etc. which deliberately juxtapose different visual codes.) As can be seen in retrospect, this new constraint eventually changed the relationship between the two paradigms in favor of a sampling paradigm. But this is only visible now, after *The Matrix* films made the sampling paradigm the basis of their visual universe.[3]

At first, when filmmakers started to incorporate synthetic 3D images in films, this did not have any effect on how computer scientists thought about computer graphics. 3D computer graphics for the first time briefly appeared in a feature film in 1980 – *Looker*. Throughout the 1980s, a number of films were made which used computer images but always only as a small element within the overall film narrative. (Released in 1982, *Tron* can be compared to *The Matrix* since its narrative universe is situated inside a computer and created through computer graphics – but this was an exception.) For instance, one of the *Star Trek* films contained a scene of a planet coming to life; it was created using the very first particle system. But this was a single scene, and it had no interaction with any other scene in the film.

In the early 1990s the situation started to change. With pioneering films such as *The Abyss* (James Cameron, 1989), *Terminator 2* (James Cameron, 1991), and *Jurassic Park* (Steven Spielberg, 1993), computer-generated characters became the key protagonists of feature films. This meant that they would appear in dozens or even hundreds of shots throughout a film, and that in most of these shots computer characters would have to be integrated with real environments and human actors captured via live action photography (called in the business 'live plate'). Examples are the T-100 cyborg character in *Terminator 2: Judgment Day*, or dinosaurs in *Jurassic Park*. These computer-generated characters are situated inside the live action universe that is the result of sampling physical reality via a 35mm film camera. The simulated world is located inside the captured world, and the two have to match perfectly.

As pointed out in *The Language of New Media* (Manovich, 2001) in the discussion of compositing, perfectly aligning elements that come from different sources is one of the fundamental challenges of computer-based realism. Throughout the 1990s, filmmakers and special effects artists have dealt with this challenge using a variety of techniques and methods. What Gaeta realized earlier than others is that the best way to align the two universes of live action and 3D computer graphics is to build *a single new universe*.[4]

Rather than treating sampling reality as just one technique to be used along with many other 'proper' algorithmic techniques of image synthesis, Gaeta and his colleagues turned it into the key foundation of the Universal Capture process. The process systematically takes physical reality apart and then systematically reassembles the elements into a virtual computer-based representation. The result is a new kind of image that has a photographic/cinematographic appearance and level of detail yet internally is structured in a completely different way.

Universal Capture was developed and refined over a three-year period from 2000 to 2003 (Borshukov, 2004). How does the process work? There are actually more stages and details involved, but the basic procedure is as follows (for more details, see Borshukov et al., 2003). An actor's performance in ambient lighting is recorded using five synchronized high-resolution video cameras. 'Performance' in this case includes everything an actor says in a film and all possible facial expressions.[5] (During production, the studio was capturing over 5 terabytes of data each day.) Next, special algorithms are used to track each pixel's movement over time at every frame. This information is combined with a 3D model of a neutral expression of the actor created using a cyberscan scanner. The result is an animated 3D shape that accurately represents the geometry of the actor's head as it changes during a particular performance. The shape is mapped with color information extracted from the captured video sequences. A separate very high resolution scan of the actor's face is used to create the map of small-scale surface details like pores and wrinkles, and this map is also added to the model.

After all the data have been extracted, aligned, and combined, the result is what Gaeta calls a 'virtual human' – a highly accurate reconstruction of the captured performance, now available as 3D computer graphics data – with all the advantages that come from having such a representation. For instance, because the actor's performance now exists as a 3D object in virtual space, the filmmaker can animate a virtual camera and 'play' the reconstructed performance from an arbitrary angle. Similarly, the virtual head also can be lighted in any way that is desired and attached to a separately constructed CG body (Borshukov et al., 2004). For example, all the characters that appeared in the Burly Brawl scene in *Matrix 2* were created by combining the heads constructed via Universal Capture done on the leading actors with CG bodies which used motion capture data from a different set of performers. Because all the characters as well as the set were computer generated, this allowed the directors of the scene to choreograph the virtual camera, making it fly around the scene in a way not possible with real cameras on a real physical set.

The process was appropriately named Total Capture because it captures all the possible information from an object or a scene using a number of recording methods – or at least, whatever is possible to capture using current technologies. Different dimensions – color, 3D geometry, reflectivity and texture – are captured separately and then put back together to create a more detailed and realistic representation.

Total Capture is significantly different from the commonly accepted methods used to create computer-based special effects such as keyframe animation and physically based modeling. In the first method, an animator specifies the key positions of a 3D model and the computer calculates in-between frames. With the second method, all the animation is automatically created by software that simulates the physics underlying the movement. (This method thus represents a particular instance of the 'reality simulation' paradigm discussed earlier.) For example, to create a realistic animation of a moving creature, the programmers model its skeleton, muscles, and skin, and specify the algorithms that simulate the actual physics involved. Often the two methods are combined: for instance, physically based modeling can be used to animate a running dinosaur while manual animation can be used for shots where the dinosaur interacts with human characters.

In recent years, the most impressive achievement in physically based modeling was the battle in *The Lord of the Rings: Return of the King* (Peter Jackson, 2003), which involved tens of thousands of virtual soldiers all driven by *Massive* software (see www.massivesoftware.com). Similar to the Non-human Players (or bots) in computer games, each virtual soldier was given the ability to 'see' the terrain and other soldiers, a set of priorities and an independent 'brain', i.e. an AI program which directs a character's actions based on perceptual inputs and priorities. However, in contrast to games AI, *Massive* software does not have to run in real time. Therefore it can create the scenes with tens and even hundreds of thousands of realistically behaving agents (one commercial created with the help of *Massive* software featured 146,000 virtual characters).

The Universal Capture method uses neither manual animation nor simulation of the underlying physics. Instead, it directly samples physical reality, including color, texture and the movement of the actors. Short sequences of the actor's performances are encoded as 3D computer animations; these animations form a library from which the filmmakers can then draw as they compose a scene. The analogy with musical sampling is obvious here. As Gaeta pointed out, his team never used manual animation to try to tweak the motion of a character's face; however, similar to a musician, they would often 'hold' a particular expression before going on to the next one (Gaeta, 2003). This suggests another analogy – editing videotape. But this is second-degree editing, so to speak: instead of simply capturing segments of reality on video and then joining them together, Gaeta's method produces complete virtual recreations of particular phenomena – self-contained micro-worlds – which can be then further edited and embedded within a larger 3D simulated space.

Animation as an idea

This brief overview of the methods of computer graphics presented here in order to explain Universal Capture offers good examples of the multiplicity of ways in which animation is used in contemporary moving image culture. If we consider this multiplicity, it is possible to come to a conclusion that 'animation' as a separate medium in fact hardly exists any more. At the same time, the general principles and techniques of putting objects and images into motion developed in 19th and 20th century animation are used much more frequently now than before computerization. But they are hardly ever used in isolation – they are usually combined with other techniques drawn from live action cinematography and computer graphics.

So where does animation start and end today? When you see a Disney animated feature or a motion graphics short it is obvious that you are seeing 'animation'. Regardless of whether the process involves drawing images by hand or using 3D software, the principle is the same: somebody created the drawings or 3D objects, set keyframes and then created inbetween positions. (Of course, in the case of commercial films, this is not just one person but large teams.) The objects can be created in multiple ways and the in-between stages can be done manually or automatically by the software, but this does not change the basic logic. Movement, or any other change over time, is defined manually – usually via keyframes (but not always). In retrospect, the definition of movement via keys was probably the essence of 20th-century animation. It was used in traditional cell animation by Disney and others, for stop motion animation by Starevich and Trnka, for the 3D animated shorts by Pixar, and it continues to be used today in animated features that combine traditional cell method and 3D computer animation. And while experimental animators such as Norman McLaren rejected keys/in-between systems in favor of drawing each frame on film by hand without explicitly defining the keys, this did not change the overall logic: the movement was created by hand. Not surprisingly, most animation artists exploited this key feature of animation in different ways, turning it into aesthetics: for instance, exaggerated squash and stretch in Disney, or the discontinuous jumps between frames in McLaren.

What about other ways in which images and objects can be set in motion? Consider, for example, the methods developed in computer graphics: physically based modeling, particle systems, formal grammars, artificial life, and behavioral animation. In all these methods, the animator does not directly create the movement. Instead it is created by the software that uses some kind of mathematical model. For instance, in the case of physically based modeling the animator may set the parameters of a computer model, which simulates a physical force such as a wind that will deform a piece of cloth over a number of frames. Or, the animator may instruct the ball to drop on the floor, and let the physics model control how the ball will bounce after it hits the floor. In the case of particle systems used to model everything from fireworks, explosions, water, and gas to animal flocks and swarms, the animator only has to define initial conditions: the number of particles, their speed, their lifespan, etc.

In contrast to live action cinema, these computer graphics methods do not capture real physical movement. Does it mean that they belong to animation? If we accept that the defining feature of traditional animation was manual creation of movement, the answer will be no. But things are not so simple. With all these methods, animators set the initial parameters, run the model, adjust the parameters, and repeat this production loop until they are satisfied with the result. So while the actual movement is produced not by hand but by a mathematical model, animators maintain significant control. In a way, animators act as film directors – only in this case they are directing not the actors but a computer model until it produces a satisfactory performance. Or we can also compare animators to film editors as they are selecting the best performances of the computer model.

James Blinn, a computer scientist responsible for creating many fundamental techniques of computer graphics, once made an interesting analogy to explain the difference between the manual keyframing method and physically based modeling.[6] He told the audience at a SIGGRAPH panel that the difference between the two methods is analogous to the difference between painting and photography. In Blinn's terms, an animator who creates movement by manually defining keyframes and drawing in-between frames is like a painter who is observing the world and then making a painting of it. The resemblance between a painting and the world depends on the painter's skills, imagination and intentions, whereas an animator who uses physically based modeling is like a photographer who captures the world as it actually is. Blinn wanted to emphasize that mathematical techniques can create a realistic simulation of movement in the physical world and an animator only has to capture what is created by the simulation. Although this analogy is useful, I think it is not completely accurate. Obviously, the traditional photographer whom Blinn had in mind (i.e. before Photoshop) chooses composition, contrast, depth of field, and many other parameters. Similarly, animators who are using physically based modeling also have control over a large number of parameters and it depends on their skills and perseverance to make the model produce a satisfying animation. Consider the following example from the related area of software art, which uses some of the same mathematical methods. Casey Reas, an artist who is well-known both for his Processing programming environment and his own still images and animations, told me recently that he may spend only a couple of hours writing a software program to create a new work – and then another two years working with the different parameters of the same program and producing endless test images until he is satisfied with the results (personal communucation, April 2005). So while, in the first instance, physically based modeling appears to be the opposite of traditional animation in that the movement is created by a computer, in fact it should be understood as a hybrid between animation and computer simulation. While animators no longer directly draw each phase of movement, they are working within the parameters of the mathematical model that 'draws' the actual movement.

And what about Universal Capture method as used in *The Matrix*? Gaeta and his colleagues also banished keyframing animation – but they did not use any mathematical modes to automatically generate motion either. As we saw, their solution was to capture the actual performances of an actor (i.e. movements of actor's face), and then reconstruct it as a 3D sequence. Together, these reconstructed sequences form a library of facial expressions. The filmmaker can then draw from this library, editing together a sequence of expressions (but not interfering with any parameters of separate sequences). It is important to stress that a 3D model has no muscles, or other controls traditionally used in animating computer graphics faces – it is used 'as is'.

Just as the case when animators employ mathematical models, this method avoids drawing individual movements by hand. And yet, its logic is that of animation rather than of cinema. The filmmaker chooses individual sequences of actors' performances, edits them, blends them if necessary, and places them in a particular order to create a scene. In short, the scene is actually constructed by hand even though its components are not. So while, in traditional animation, the animator draws each frame to create a short sequence (for instance, a character turning his head), here the filmmaker 'draws' on a higher level, manipulating whole sequences as opposed to their individual frames.

To create final movie scenes, Universal Capture is combined with Virtual Cinematography, staging the lighting, the positions and movement of a virtual camera that is 'filming' the virtual performances. What makes this Virtual Cinematography as opposed to simply computer graphics? The reason is that the world as seen by a virtual camera is different from the normal world of computer graphics. It consists of reconstructions of the actual set and the actual performers created via Universal Capture. The aim is to avoid the more manual processes usually used to create 3D models and sets. Instead, the data relating to the physical world are captured and then used to create a precise virtual replica.

Ultimately, ESC's production method as used in *The Matrix* is neither 'pure' animation, nor cinematography, nor traditional special effects, nor traditional computer animation. And this is typical of moving image culture today. When the techniques drawn from these different traditions are fused together in a computer environment, the result is not a sum of components but a variety of hybrid methods such as Universal Capture. I believe that this is how different moving image techniques function now in general. After computerization virtualizes them – 'extracting' them from their particular physical media to turn them into algorithms – they start interacting and creating hybrids. Which means that, in most cases, we will no longer find any of these techniques in their pure original state.

For instance, what does it mean when we see depth of field effect in motion graphics, films and television programs which use neither live action footage nor photorealistic 3D graphics but have a more stylized look? Originally an artifact of lens-based recording, depth of field was simulated in a computer when the main goal of the 3D computer graphics field was to create maximum 'photorealism', i.e. synthetic scenes indistinguishable from live action cinematography.[7] But once this technique became available, moving image artists gradually realized that it could be used regardless of how realistic or abstract the visual style is – as long as there is a suggestion of a 3D space. Typography moving in perspective through an empty space, drawn 2D characters positioned on different layers in a 3D space, a field of animated particles – any composition can be put through the simulated depth of field.

The fact that this effect is simulated and removed from its original physical media means that a designer can manipulate it in a variety of ways. The parameters that define what part of the space is in focus can be independently animated, i.e. set to change over time, because they are simply the numbers controlling the algorithm and not something built into the optics of a physical lens. So while simulated depth of field can be said to maintain the memory of the particular physical media (lens-based photo and film recording) from which it came, it became an essentially new technique that functions as a 'character' in its own right. It has the fluidity and versatility not previously available. Its connection to the physical world is ambiguous at best. On the one hand, it only makes sense to use depth of field if you are constructing a 3D space – even if it is defined in a minimal way by using only a few or even a single depth cue, such as lines converging towards the vanishing point or foreshortening. On the other hand, the designer can be said to 'draw' this effect in any way desired. The axis controlling depth of field does not need to be perpendicular to the image plane, the area in focus can be anywhere in space, it can also quickly move around the space, etc.

Coming back to Universal Capture, it is worthwhile to quote Gaeta who himself is very clear that what he and his colleagues have created is a new hybrid. In an interview in 2004, he said: 'If I had to define virtual cinema, I would say it is somewhere between a live-action film and a computer-generated animated film. It is computer generated, but it is derived from real world people, places and things' (Feeny, 2004). Although Universal Capture offers a particularly striking example of Gaeta's 'somewhere between', most forms of moving image created today are similarly 'somewhere between', with animation being one of the coordinate axes of this new space of hybridity.

'Universal Capture': reality re-assembled

The method which came to be called 'Universal Capture' combines the best of two worlds: physical reality as captured by lens-based cameras, and synthetic 3D computer graphics. While it is possible to recreate the richness of the visible world through manual painting and animation, as well as through various computer graphics techniques (texture mapping, bump mapping, physical modeling, etc.), it is expensive in terms of the labor involved. Even with physically based modeling techniques, endless parameters have to be tweaked before the animation looks right. In contrast, capturing visible reality through the lens on film, tape, DVD-R, computer hard drive, or other media is cheap: just point the camera and press the 'record' button.

The disadvantage of such recordings is that they lack the flexibility demanded by the contemporary remix culture. This culture demands not self-contained aesthetic objects or self-contained records of reality but smaller units – parts that can be easily changed and combined with other parts in endless combinations. However, a lens-based recording process flattens the semantic structure of reality – i.e. the different objects that occupy distinct areas of a 3D physical space. It converts a space filled with discrete objects into a flat field of image grains or pixels that do not carry any information of where they came from (i.e. which objects they correspond to). Therefore, any kind of editing operation – deleting objects, adding new ones, compositing, etc. – becomes quite difficult. Before anything can be done with an object in the image, it has to be manually separated by creating a mask. And unless an image shows an object that is properly lighted and shot against a special blue or green background, it is impossible to mask the object precisely.

In contrast, 3D computer generated worlds have the exact flexibility one would expect from media in the information age. (It is therefore not surprising that 3D computer graphics representation – along with hypertext and other new computer-based data representation methods – was conceptualized in the same decade that the transformation of advanced industrialized societies into information societies became apparent.) In 3D computer-generated worlds, everything is discrete. The world consists of a number of separate objects. Objects are defined by points described as XYZ coordinates; other properties of objects such as color, transparency and reflectivity are similarly described in terms of discrete numbers. This means that the semantic structure of a scene is completely preserved and is easily accessible at any time. To duplicate an object a hundred times requires only a few mouse clicks or typing in a short command; similarly, all other properties of a world can always be easily changed. And since each object itself is made up of discrete components (flat polygons or surface patches defined by splines), it is equally easy to change its 3D form by selecting and manipulating its components. In addition, just as a sequence of genes contains the code that is expanded into a complex organism, a compact description of a 3D world that contains only the coordinates of the objects can be quickly transmitted through the network, with the client computer reconstructing the full world (this is how online multiplayer computer games and simulators work).

Starting in the late 1970s when James Blinn (1978) introduced texture mapping, computer scientists, designers and animators gradually expanded the range of information that could be recorded in the real world and then incorporated it into a computer model. Until the early 1990s, this information mostly involved the appearance of objects: color, texture, light effects. The next significant step was the development of motion capture. During the first half of the 1990s, it was quickly adopted in the movie and game industries. Now computer-synthesized worlds relied not only on sampling the visual appearance of the real world but also on sampling movements of animals and humans in this world. Building on all these techniques, Gaeta's method takes them to a new stage: capturing just about everything that at present can be captured, and then reassembling the samples to create a digital (and thus completely malleable) recreation. Put in a larger context, the resulting 2D/3D hybrid representation perfectly fits with the most progressive trends in contemporary culture which are all based on the idea of a hybrid.

The new hybrid

It is my strong feeling that the emerging 'information aesthetics' (i.e. the new cultural features specific to information society) has or will have a very different logic from modernism. The latter was driven by a strong desire to erase the old – visible as much in the avant-garde artists' (particularly the Futurists') statements that museums should be burned, as in the dramatic destruction of all social and spiritual realities of many people in Russia after the 1917 revolution, and in other countries when they became Soviet satellites after 1945. Culturally and ideologically, modernists wanted to start with a *tabula rasa*, radically distancing them from the past. It was only in the 1960s that this move started to feel inappropriate, as manifested both in the loosening ideology in the communist countries and the beginnings of a new postmodern sensibility in the West. To quote the title of a famous book written by Robert Venturi et al. (1977[1972]), *Learning from Las Vegas* (the first systematic manifestation of a new sensibility) meant admitting that organically developing vernacular cultures involve bricolage and hybridity rather than purity – seen, for instance, in 'international style', which was still practised by architects worldwide at that time. Driven less by the desire to imitate vernacular cultures and more by the new availability of previous cultural artifacts stored on magnetic and later digital media, commercial culture in the West in the 1980s systematically replaced purity with stylistic heterogeneity. Finally, when the Soviet Empire collapsed, postmodernism had won the world over.

Today we have a very real danger of being imprisoned by a new 'international style' – something that can be called 'global international'. Cultural globalization, of which cheap airline flights and the internet are the two most visible representatives, erases a certain cultural specificity with the energy and speed that modernism could not emulate. Yet we are also witnessing today a different logic at work: the desire to creatively place together the old and the new – local and transnational – in various combinations. It is this logic, for instance, which has made a city such as Barcelona (where I talked with John Gaeta in the context of the Art Futura 2003 festival which led to this article) such a 'hip' and 'in' place today. All over Barcelona, architectural styles of many past centuries co-exist with new 'cool' spaces of bars, hotels, museums, and so on. Medieval meets multinational, Gaudi meets Dolce and Gabbana, Mediterranean time meets internet time. The result is the incredible sense of energy which one feels physically just walking along the street. It is this hybrid energy that characterizes in my view the most interesting cultural phenomena today.[8] The hybrid 2D/3D image of *The Matrix* is one such example.

The historians of cinema often draw a contrast between the Lumières and Marey. Along with a number of inventors in other countries all working independently from each other, the Lumières created what we now know as cinema with its visual effect of continuous motion based on the perceptual synthesis of discrete images. Earlier, Muybridge had already developed a way to take successive photographs of a moving object such as a horse; eventually the Lumières and others figured out how to take enough samples so that, when projected, they perceptually fuse into continuous motion. Being a scientist, Marey was driven by an opposite desire: not to create a seamless illusion of the visible world but rather to be able to understand its structure by keeping subsequent samples discrete. Since he wanted to be able to easily compare these samples, he perfected a method where the subsequent images of moving objects were superimposed within a single image, thus making the changes clearly visible.

The hybrid image of *The Matrix* in some ways can be understood as the synthesis of these two approaches, which for a hundred years remained in opposition. Like the Lumières, Gaeta's goal is to create a seamless illusion of continuous motion. At the same time, like Marey, he also wants to be able to edit and sequence the individual recordings.

At the beginning of this article I evoked the notion of uneven development, pointing out that often the inside structure ('infrastructure') completely changes before the surface ('superstructure') catches up. What does this idea imply for the future of images and in particular 2D/3D hybrids as developed by Gaeta and others? As Gaeta (2003) pointed out, while his method can be used to make all kinds of images, so far it has been used in the service of realism as defined in cinema, i.e. anything the viewer sees has to obey the laws of physics (Gaeta, 2003). So, in the case of *The Matrix*, its images still have a traditional 'realistic' appearance while internally they are structured in a completely new way. In short, we see the old 'superstructure' that still sits on top of the 'new' infrastructure. What kinds of images will we see when the superstructure finally catches up with the infrastructure?

Of course, although current images of Hollywood special effects movies have so far followed the constraint of realism, i.e. obeying the laws of physics, they are also not exactly the same as before. In order to sell movie tickets, DVDs, and all other merchandise, each new special effects film tries to top the previous one, showing something that nobody has seen before. In *The Matrix* it was 'bullet time'; in *Matrix 2* it was the Burly Brawl scene where dozens of identical clones fight Neo; in *Matrix 3* it was the Superpunch (Borshukov, 2004). The fact that the internal construction of the image is different allows for all kinds of new effects; listening to Gaeta, it is clear that for him the key advantage of such images is the possibilities they offer for virtual cinematography. That is, if previously camera movement was limited to a small and well-defined set of moves – pan, dolly, roll – now it can move in any trajectory imaginable for as long as the director wants. Gaeta talks about the Burly Brawl scene in terms of *virtual choreography*: this implies choreographing the intricate and long camera moves that would be impossible in the real word as well as all the bodies participating in the fight (all of them are digital recreations assembled using Gaeta's method described earlier).

According to Gaeta, creating just this one scene took about three years. So while, in principle, Gaeta's method represents the most flexible way to recreate visible reality in a computer so far, it will be years before this method is sufficiently streamlined and standardized for these advantages to become obvious. But when it happens, the artists will have an extremely flexible hybrid medium at their disposal: completely virtualized cinema. Rather than expecting that any of the present pure forms will dominate the future of visual culture, I think this future belongs to such hybrids. In other words, future images will probably still be photographic – although only on the surface.

And what about animation? What will be its future? As I have tried to explain, besides purely animated films and animated sequences used as a part of other moving image projects, animation has become a set of principles and techniques that animators and filmmakers employ today to create new methods and new visual styles. Therefore, I think it is not worth asking if this or that visual style or method for creating moving images that emerged after computerization is 'animation' or not. It is more constructive to say that most of these methods were born from animation and have animation DNA – mixed with DNA from other media. I think that such a perspective that considers 'animation in an extended field' is a more productive way to think about animation today, especially if we want our reflections to be relevant for everybody concerned with contemporary visual and media cultures.

Notes

[1] For technical details of the method, see the publications of Georgi Borshukov [www.virtualcinematography.org/publications.html].

[2] Although not everybody would agree with this analysis, I feel that after the end of the 1980s, the field significantly slowed down: on the other hand, all key techniques that can be used to create photorealistic 3D images have already been discovered. The rapid development of computer hardware in the 1990s meant that computer scientists no longer had to develop new techniques to make the rendering faster, since the already developed algorithms would now run fast enough.

[3] The terms 'reality simulation' and 'reality sampling' have been invented for this article; the terms 'virtual cinema', 'virtual human', 'universal capture' and 'virtual cinematography' come from John Gaeta. The term 'image-based rendering' first appeared in the 1990s.

[4] Therefore, while the article in *Wired* which positioned Gaeta as a groundbreaking pioneer and as a rebel working outside Hollywood contained the typical journalistic exaggeration, it was not that far from the truth (Silberman, 2003).

[5] The method captures only the geometry and images of an actor's head; body movements are recorded separately using motion capture.

[6] I am not sure about the exact year of the SIGGRAPH conference where Blinn gave his presentation, but I think it was the end of the 1980s when physically based modeling was still a new concept.

[7] For more on this process, see the chapter 'Synthetic Realism and its Discontents' in Manovich (2001).

[8] Seen from this perspective, my earlier book *The Language of New Media* (2001) can be seen as a systematic investigation of a particular slice of contemporary culture driven by this hybrid aesthetics: the slice where the logic of digital networked computer intersects the numerous logics of already established cultural forms.

References

Blinn, J.F. (1978) ‘Simulation of Wrinkled Surfaces’, *Computer Graphics*, August: 286–92.

Borshukov, Georgi (2004) ‘Making of the Superpunch’, presentation at Imagina 2004, available at [ww.virtualcinematography.org/publications/acrobat/Superpunch.pdf].

Borshukov, Georgi, Piponi, Dan, Larsen, Oystein, Lewis, J.P. and Tempelaar-Lietz, Christina (2003) ‘Universal Capture – Image-Based Facial Animation for “The Matrix Reloaded”’, SIGGRAPH 2003 Sketches and Applications Program, available at [http://www.virtualcinematography.org/publications/acrobat/UCap-s2003.pdf].

Feeny, Catherine (2004) ‘“The Matrix” Revealed: An Interview with John Gaeta’, VFXPro, 9 May [www.uemedia.net/CPC/vfxpro/article_7062.shtml]

Gaeta, John (2003) Presentation during a workshop on the making of *The Matrix*, Art Futura 2003 festival, Barcelona, 12 October.

Manovich, Lev (2001) *The Language of New Media*. Cambridge, MA: MIT Press.

Silberman, Steve (2003) ‘Matrix 2’, *Wired*, 11 May [http://www.wired.com/wired/archive/11.05/matrix2.html]

Venturi, Robert, Izenour, Steven and Scott Brown, Denise (1977[1972]) *Learning from Las Vegas: The Forgotten Symbol of Architectural Form*, rev edn. Cambridge, MA: MIT Press.

Lev Manovich is Professor of New Media, Visual Arts Department, and Director of The Lab for Cultural Analysis, CAL-IT2, at the University of California, San Diego.

Address: University of California, San Diego, Visual Arts Department, 9500 Gilman Drive MC 0084, La Jolla, CA 92093–0084, USA. [email: lev@manovich.net]