# Image Future

*author: Lev Manovich*

*year: 2006*

## Introduction

For the larger part of the twentieth 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. Softwarization of all areas of moving image production created a common pool of techniques that can be used regardless of whether one is creating motion graphics for television, a narrative feature, an animated feature, or a music video. The abilities to composite many layers of imagery with varied transparency, to place 2D and 3D visual 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 of a frame are equally available to the creators of all forms of moving images.

The existence of this common vocabulary of software-based techniques does not mean that all films now look the same. What it means, however, is that while most live action films, animated features and motion graphics do look quite distinct today, this is the result of a deliberate choices rather than the inevitable consequence of differences in production methods and technology.

Given that all techniques of previously distinct media are now available within a single software-based production environment, what is the meaning of the terms that were used to refer to these media in the twentieth century – such as "animation"? From the industry point of view, the answer is simple. Animation not only continues to exist as a distinct area of media industry but it's also very successful – its success in no small part fueled by new efficiency of software-based global production workflow. 2D and 3D animated features,

shorts and series are produced today in larger numbers than ever before; students can pursue careers in "animation"; Japanese anime and animated features continue to grow in popularity; China is building whole cities around mega-size animation and rendering studios and production facilities.

Certainly, the aesthetics of many contemporary feature-length 3D animated features largely relies on the visual language of twentieth-century commercial animation. So, while everything may be modeled and animated in 3D computer animation program, the appearance of the characters, their movements, and the staging of scenes conceptually owe more to mid-20th century Disney than to 21st century Autodesk (producer of industry-standard Maya software). Similarly, hybrid looking short-form films (exemplified by but not limited to "motion graphics") also often feature sequences or layers that look very much like character animation we know from the 20th century.

The examples above illustrate just one, more obvious, role of animation in contemporary post-digital visual landscape. In this chapter I will explore its other role: as a generalized tool set 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. Put differently, what has been "animation" has become a part of the computer metamedium.

I have chosen a particular example for my discussion that I think will illustrate well this new role of animation. It is an especially intricate method of combining live action and CG (a common abbreviation for "computer graphics.") Called "Universal Capture" (U-cap) by their creators, it was first systematically used on a large scale by ESC Entertainment in *Matrix 2* and *Matrix 3* 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. The use of Universal Capture also leads to visual hybrids – but they are quite distinct from the hybrids found in motion graphics and other short-form moving image productions being created today. With Universal Capture, different types of imagery are "fused" together to create a new kind of image. This image combines "the best of" qualities of two types of imagery that we normally understand as being ontologically

the 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 gradually diminish.

## Uneven Development

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

There are good reasons to assume that the future images would 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. One of the reason for this persistence of 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 continue to dominate contemporary visual culture, most of them are not pure photographs but various mutations and hybrids: photographs which went through different 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 which are not limited to the part of the spectrum visible to a human eye (night vision, x-ray); simulated photographs created 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 Marxist theory of historical development, infrastructure (i.e., mode of production in a given society – also called "base") changes well before superstructure (i.e., ideology and culture in this society). To use a different example, think of the history of technology in the twentieth century. Typically, a new type of machine was at first fitted within old, familiar skin: for instance, early twentieth century cars emulated the form of horse carriage. The popular idea usually ascribed to Marshall McLuhan – that the new media first emulates 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 old structure of media organization, before the new structure emerges. For instance, first typeset book was designed to emulate hand-written books; cinema first emulated theatre; and so on.

This concept of uneven development can be useful in thinking about the changes in contemporary visual culture. Since this process started in the middle of the 1950s, 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, still largely remains the same. It is therefore possible that at some point in the future the "skin" of a photographic image would also become completely different, but this did not happen yet. So, we can say at present our visual culture is characterized by a new computer "base" and old photographic "superstructure."

*The Matrix* films provide us with a very rich set of examples perfect for thinking further about these issues. The trilogy is an allegory about how its visual universe is constructed. That is, the films tell us about the Matrix, a virtual universe that is maintained by computers – and of course, visually the images of *The Matrix* trilogy that we the viewers see in the films were all indeed assembled using help 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 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 the futuristic vision of computer games in the future when it would become 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 Paul Debevec, Georgi Borshukov, John Gaeta, and a number of other people both in academia and in the special effects industry. [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*, 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 special effects in *The Matrix* rely on Universal Capture. Also, since *The Matrix*, other Hollywood films and video games (EA SPORT Tiger Woods 2007) already used some of the same strategies. However, in this chapter I decided to focus on the use of this process in the second and third films of *The Matrix* for which the method of Universal Capture was originally developed. And while the complete credits for everybody involved in developing Universal Capture would run for a whole page, here I will identify it with Gaeta. The reason is not because, as a senior special effects supervisor for *The Matrix Reloaded* and *The Matrix Revolutions* 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, lets briefly run through the history of 3D photo-realistic image synthesis and its use in the film industry. In 1963 Lawrence G. Roberts (who later in the 1960s became one of the key people behind the development of Arpanet but at that time was a graduate student at MIT) 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 would be called "wire frames." Approximately ten 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 create 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 combination of these techniques to synthesize images of almost every subject that often were not easily distinguishable from traditional cinematography. ("Almost" is important here since the creation of photorealistic moving images of human faces remained a hard to reach a goal – and this is in part what Total Capture method was designed to address.)

All this research was based on one fundamental assumption: in order to re-create an image of visible 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 this physics, a computer would take forever to calculate even a single image, they put their energy in inventing various short cuts which 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 is one such "hack" – a particular approximation of a particular subset of all possible interactions between light sources, materials, and cameras.

This assumption also meant that you are re-creating reality step-by-step starting from a blank canvas (or, more precisely, an empty 3D space.) 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 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 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 sights 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 the 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 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 have 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 already 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 be also anything else, for instance a logo, a photograph of a face or of clouds – is wrapped around a 3D model. This is a very effective way to add 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 which make up the 3D texture of a concrete wall, an artist can simply take a photograph of an existing wall, convert into a grayscale image, and then feed this image to the rendering algorithm. The algorithm treats grayscale image as a depth map, i.e., the value of every pixel is being interpreted as 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 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 weaved 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 have been always widely used as soon as they were invented, many people in the computer graphics field always felt that they were cheating. Why? I think this feeling was there 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 rarely would fit exactly right, and painstaking manual adjustments were required. For instance, texture mapping would work perfectly if applied to a flat 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 above, sampling paradigm was "imbedded" within reality simulation paradigm. It was a 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 and then adding these samples to a virtual scene was a trick, a shortcut within over wise honest game of mathematically simulating reality in a computer.

## Building The Matrix

So far, we looked at the paradigms of 3D computer graphics field without considering the uses of the 3D images? So, what happens if you want to incorporate photorealistic images produced with CG 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 juxtaposes different visual codes). As can be seen in retrospect, this new constraint eventually changed the relationship between the two paradigms in favor of sampling paradigm. But this is only visible now, after films such as *The Matrix* made the sampling paradigm the basis of its 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. (One exception was *Tron*; released in 1982, it can be compared to *The Matrix* since its narrative universe is situated inside computer and created through computer graphics – but this was an exception.) For instance, one of *Star Trek* films contained a scene of a planet coming to life; it was created using CG. (In fact, now commonly used "particle system" was invented for to crate this effect.) But this was a single scene, and it had no interaction with all other scenes in the film.

In the early 1990s the situation has 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

(such shots are called in the business "live plates.") 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 capturing physical reality via the lens of a film camera. The simulated world is located inside the captured world, and the two have to match perfectly.

As I pointed out in *The Language of New Media* in the discussion of compositing, perfectly aligning elements that come from different sources is one of 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 was 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 Universal Capture process. The process systematically takes physical reality apart and then systematically reassembles the elements together to create a new software-based representation. The result is a new kind of image that has 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. [5] How does the process work? There are actually more stages and details involved, but the basic procedure is the following. [6] An actor's performance is recorded using five synchronized high-resolution video cameras. "Performance" in this case includes everything an actor will say in a film and all possible facial expressions. [7] (During the 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 captured via a 3D 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. (How is that for hybridity?)

After all the data has been extracted, aligned, and combine, the result is what Gaeta calls a "virtual human" - a highly accurate reconstruction of the captured performance, now available as a 3D computer graphics data – with all the advantages that come from having such representation. For instance, because actor's performance now exists as a 3D object in virtual space, the filmmaker can animate virtual camera and "play" the reconstructed performance from an arbitrary angle. Similarly, the virtual head can be also lighted in any way desirable. It can be also attached to a separately constructed CG body. [8] For example, all the characters which appeared the Burly Brawl scene in *The 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 along with the set were computer generated, this allowed the directors of the scene to choreograph the virtual camera, having 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 "reality simulation" paradigm.) For instance, to create a realistic animation of 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.

When the third *Matrix* film was being released, 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. [9] 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., a AI program which directs character's actions based on the perceptual inputs and priorities. But 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 realistically behaving agents (one commercial created with the help of Massive software featured 146,000 virtual characters).

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 an 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 character's face; however, just as a musician may do it, they would often "hold" particular expression before going to the next one. [10] This suggests another analogy – analog video editing. But this is a 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

The brief overview of the methods of computer graphics that I presented above 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 anymore. At the same time, the general principles and techniques of putting objects and images into motion developed in nineteenth and twentieth century animation are used much more frequently now than before computerization. But they are hardly ever used by themselves – usually they are 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 or Pixar animated feature, or many graphics shorts 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 in-between positions. (Of course, in the course of commercial films, this is not one person but large teams.) The objects can be created in multiple ways and inbetweening can be done manually or automatically by the software, but this does not change the basic logic. The movement, or any other change over time, is defined manually – usually via keyframes (but not always). In retrospect, the definition of movement via keys probably was the essence of twentieth 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 refused keys / in-betweens system 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 into 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 sets the parameters of a computer model which simulates a physical force such as wind which will deform a piece of cloth over a number of frames. Or she 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: a 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, the animator sets the initial parameters, runs the model, adjusts the parameters, and repeats this production loop until she is satisfied with the result. So, while the actual movement is produced not by hand by a mathematical model, the animator maintains significant control. In a way, the animator acts as a film director – only in this case she is directing not the actors but the computer model until it produces a satisfactory performance. Or we can also compare her to a film editor who is selecting among 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 manual keyframing method and physically based modeling. [11] 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 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, an animator who is using physically based modeling also has control over a large number of parameters and it depends on her skills and perseverance to make the model produce a satisfying animation. Consider the following example from the related area of software art that uses some of the same mathematical methods. Casey Reas, an artist who is well-known both for his own still images and animations and for Processing graphics programming environment he helped to develop, told me 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. [12] So while at first physically based modeling appears to be 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 the animator no longer directly draws each phase of movement, she is working with 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 used 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 it is the case when animator employs 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 animation" as we already know it? The reason is that the world as seen by a virtual camera is different from a 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 manual processes usually used to create 3D models and sets. Instead, the data about the physical world is 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 CG. Instead, it is "pure" example of hybridity in general, and "deep remixability" in particular. With its complex blend of the variety of media techniques and media formats, it is also typical of moving image culture today. When the techniques drawn from these different media traditions are brought together in a software environment, the result is not a sum of separate components but a variety of hybrid methods - such as Universal Capture. As I already noted more than once, I think that this how different moving image techniques function now in general. After computerization virtualized them – "extracting" them from their particular physical media to turn into algorithms – they start interacting and creating hybrids. While we have already encountered various examples of hybrid techniques, Total Capture and Virtual Cinematography illustrate how creative industries today develop whole production workflow based on hybridity.

It is worthwhile here to quote Gaeta who himself is very clear that what he and his colleagues have created is a new hybrid. In 2004 interview, he says: "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." [13] Although Universal Capture offers a particularly striking example of such "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: visible 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 labor involved. Even with physically based modeling techniques endless parameters have to be tweaked before the animation looks right. In contrast, capturing visible reality via lens-based recording (the process which in the twentieth century was called "filming") is cheap: just point the camera and press "record" button.

The disadvantage of such lens-based recordings is that they lack flexibility demanded by contemporary remix culture. *Remix 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, lens-based recording process flattens the semantic structure of reality. Instead of a set of unique objects which occupy distinct areas of a 3D physical space, we end up with a flat field of made from pixels (or film grains in the case of film-based capture) that do not carry any information of where they came from, i.e., which objects they correspond to. Therefore, any kind of spatial 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 from the rest of the image 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 practically impossible to mask the object precisely.

In contrast, 3D computer generated worlds have the exact flexibility one would expect from media in information age. (It is not therefore accidental that 3D computer graphics representation – along with hypertext and other new computer-based data representation methods – was conceptualized in the same decade when the transformation of advanced industrialized societies into information societies became visible.) In a 3D computer generated worlds everything is discrete. The world consists of a number of separate objects. Objects are defined by points described by their coordinates in a 3D space; other properties of objects such as color, transparency and reflectivity are similarly described in terms of discrete numbers. As a result, while a 3D CG representation may not have the richness of a lens-based recording, it does contain a semantic structure of the world. This structure is easily accessible at any time. A designer can directly select any object (or any object part) in the scene. Thus, to duplicate an object hundred times requires only a few mouse clicks or typing a short command; similarly, all other properties of a world can be always easily changed. And since each object itself consists 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 multi-player computer games and simulators work).

Universal Capture brings together the complementary advantages of lens-based capture and CG representation in an ingenious way. Beginning in the late 1970s when James Blinn introduced CG technique of texture mapping [14], computer scientists, designers and animators were gradually expanding the range of information that can be recorded in the real world and then incorporated into a computer model. Until the early 1990s this information mostly involved the appearance of the 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 of 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

18

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) already has or will have a very different logic from what modernism. The later 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 well as in the dramatic destruction of all social and spiritual realities of many people in Russia after the 1917 revolution, and in other countries after they became Soviet satellites after 1945. Culturally and ideologically, modernists wanted to start with "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 loosening of ideology in communist countries and the beginnings of new post-modern sensibility in the West. To quote the title of a famous book by Robert Venturi, Denise Scott Brown, and Steven Izenour (published in 1972, it was the first systematic manifestation of new sensibility), *Learning from Las Vegas* meant admitting that organically developing vernacular cultures involves bricolage and hybridity, rather than purity seen for instance in "international style" which was still practiced by architects world-wide 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 soon digital media, in the 1980s commercial culture in the West systematically replaced purity by stylistic heterogeneity. Finally, when Soviet Empire collapsed, post-modernism has won world over.

Today we have a very real danger of being imprisoned by new "international style" - something which we can call the new "global style" The cultural globalization, of which cheap airline flights, the web, and billions of mobile phones are two most visible carriers, erases some dimensions of the cultural specificity with the energy and speed impossible for modernism. Yet we also witness today a different logic at work: the desire to creatively

place together old and new – local and transnational - in various combinations. It is this logic, for instance, which made cities such as Barcelona (where I talked with John Gaeta in the context of Art Futura 2003 festival which led to this article), such a "hip" and "in" place at the turn of the century (that is, 20th to 21st). All over Barcelona, architectural styles of many past centuries co-exist with new "cool" spaces of bars, lounges, hotels, new museums, and so on. Medieval meets multi-national, Gaudi meets Dolce and Gabbana, Mediterranean time meets global time. The result is the invigorating sense of energy which one feels physically just walking along the street. It is this hybrid energy, which characterizes in my view the most interesting cultural phenomena today. [15] The hybrid 2D / 3D image of *The Matrix* is one such hybrids.

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 already developed a way to take successive photographs of a moving object such as horse; eventually the Lumières and others figured out how to take enough samples so 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 understand as the synthesis of these two approaches which for a hundred years ago 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 of reality.

In the beginning of this chapter, I evoked the notion of uneven development, pointing that often the structure inside ("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 pointed out in 2003, while his method can be used to make all kinds of images, so far it was used in the service of realism as it is defined in cinema – i.e., anything the viewer will see has to obey the laws of physics. [16] So in the case of *The Matrix,* its images still have traditional "realistic" appearance while internally they are structured in a completely new way. In short, we see the old "superstructure" which stills sits on top of "old" infrastructure. What kinds of images would we see then the superstructure" would finally catch up with the infrastructure?

Of course, while the images of Hollywood special effects movies so far follow the constraint of realism, i.e., obeying the laws of physics, they are also continuously expanding the boundaries of what "realism" means. 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* 1 it was "bullet time"; in *The Matrix* 2 it was the Burly Brawl scene where dozens of identical clones fight Neo; in *Matrix 3* it was the Superpunch. [17] The fact that the image is constructed differently internally does allow for all kinds of new effects; listening to Gaeta it is clear that for him the key advantage of such image is the possibilities it offers for virtual cinematography. That is, if before 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*: both choreographing the intricate and long camera moves impossible in the real word and also all the bodies participating in the flight (all of them are digital recreations assembled using Total Capture method). According to Gaeta, creating this one scene took about three years. So, while in principle Total Capture represents one of the most flexible way to recreate visible reality in a computer so far, it will be years before this method is streamlined and standardized enough 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, the future images would probably be still photographic – although only on the surface.

And what about animation? What will be its future? As I have tried to explain, besides animated films proper and animated sequences used as a part of other moving image projects, animation has become a set of principles and techniques which animators, filmmakers and designers employ today to create new techniques, new production methods and new visual aesthetics. Therefore, I think that it is not worthwhile to ask if this or that visual style or method for creating moving images which emerged after computerization is "animation" or not. It is more productive 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 which considers "animation in an extended field" is a more productive way to think about animation today, and that it also applies to other modern media fields which "donated" their genes to a computer metamedium.

## References:

[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 1980s the field has significantly slowed down: on the other hand, all key techniques which can be used to create photorealistic 3D images have been already discovered; on the other hand, 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" are made up by me; the terms "virtual cinema," "virtual human," "universal capture" and "virtual cinematography" come from John Gaeta. The term "image based rendering" appeared already in the 1990s – see the publication list at http://www.debevec.org/Publications/, accessed September 4, 2008.

[4] Therefore, while the article in *Wired* which positioned Gaeta as a groundbreaking pioneer and as a rebel working outside of Hollywood contained the typical journalistic

exaggeration, it was not that far from the truth. Steve Silberman, "Matrix 2," *Wired* 11.05 (May 2003), http://www.wired.com/wired/archive/11.05/matrix2.html

[5] Georgi Borshukov, "Making of *The Superpunch*," presentation at Imagina 2004, available at www.virtualcinematography.org/publications/acrobat/Superpunch.pdf.

[6] The details can be found in George Borshukov, Dan Piponi, Oystein Larsen, J.P.Lewis, Christina Tempelaar-Lietz, "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.

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

[8] Borshukov et al, "Universal Capture."

[9] See www.massivesoftware.com.

[10] John Gaeta, presentation during a workshop on the making of *The Matrix*, Art Futura 2003 festival, Barcelona, October 12, 2003.

[11] I don't remember the exact year of SIGGRAPH conference where Blinn has spoken but I think it was end of the 1980s when physically based modeling was still a new concept.

[12] Casey Reas, private communication, April 2005.

[13] Catherine Feeny, "_The Matrix_ Revealed: An Interview with John Gaeta," *VFXPro*, May 9, 2004, www.uemedia.net/CPC/vfxpro/article_7062.shtml.

[14] J. F. Blinn, "Simulation of Wrinkled Surfaces," *Computer Graphics* (August 1978): 286-92.

[15] Seen in this perspective, my earlier book *The Language of New Media* 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. Lev Manovich, *The Language of New Media* (The MIT Press, 2001.)

[16] John Gaeta, making of *The Matrix* workshop.

[17] Borshukov, "Making of *The Superpunch*."