



November 23, 2008

IDEA LAB

Becoming Screen Literate

By KEVIN KELLY

Everywhere we look, we see screens. The other day I watched clips from a movie as I pumped gas into my car. The other night I saw a movie on the backseat of a plane. We will watch anywhere. Screens playing video pop up in the most unexpected places — like A.T.M. machines and supermarket checkout lines and tiny phones; some movie fans watch entire films in between calls. These ever-present screens have created an audience for very short moving pictures, as brief as three minutes, while cheap digital creation tools have empowered a new generation of filmmakers, who are rapidly filling up those screens. We are headed toward screen ubiquity.

When technology shifts, it bends the culture. Once, long ago, culture revolved around the spoken word. The oral skills of memorization, recitation and rhetoric instilled in societies a reverence for the past, the ambiguous, the ornate and the subjective. Then, about 500 years ago, orality was overthrown by technology. Gutenberg's invention of metallic movable type elevated writing into a central position in the culture. By the means of cheap and perfect copies, text became the engine of change and the foundation of stability. From printing came journalism, science and the mathematics of libraries and law. The distribution-and-display device that we call printing instilled in society a reverence for precision (of black ink on white paper), an appreciation for linear logic (in a sentence), a passion for objectivity (of printed fact) and an allegiance to authority (via authors), whose truth was as fixed and final as a book. In the West, we became people of the book.

Now invention is again overthrowing the dominant media. A new distribution-and-display technology is nudging the book aside and catapulting images, and especially moving images, to the center of the culture. We are becoming people of the screen. The fluid and fleeting symbols on a screen pull us away from the classical notions of monumental authors and authority. On the screen, the subjective again trumps the objective. The past is a rush of data streams cut and rearranged into a new mashup, while truth is something you assemble yourself on your own screen as you jump from link to link. We are now in the middle of a second Gutenberg shift — from book fluency to screen fluency, from literacy to visuality.

The overthrow of the book would have happened long ago but for the great user asymmetry inherent in all media. It is easier to read a book than to write one; easier to listen to a song than to compose one; easier to attend a play than to produce one. But movies in particular suffer from this user asymmetry. The intensely collaborative work needed to coddle chemically treated film and paste together its strips into movies meant that it was vastly easier to watch a movie than to make one. A Hollywood blockbuster can take a million person-hours to produce and only two hours to consume. But now, cheap and universal tools of creation (megapixel phone cameras, Photoshop, iMovie) are quickly reducing the effort needed to create moving

images. To the utter bafflement of the experts who confidently claimed that viewers would never rise from their reclining passivity, tens of millions of people have in recent years spent uncountable hours making movies of their own design. Having a ready and reachable audience of potential millions helps, as does the choice of multiple modes in which to create. Because of new consumer gadgets, community training, peer encouragement and fiendishly clever software, the ease of making video now approaches the ease of writing.

This is not how Hollywood makes films, of course. A blockbuster film is a gigantic creature custom-built by hand. Like a Siberian tiger, it demands our attention — but it is also very rare. In 2007, 600 feature films were released in the United States, or about 1,200 hours of moving images. As a percentage of the hundreds of millions of hours of moving images produced annually today, 1,200 hours is tiny. It is a rounding error.

We tend to think the tiger represents the animal kingdom, but in truth, a grasshopper is a truer statistical example of an animal. The handcrafted Hollywood film won't go away, but if we want to see the future of motion pictures, we need to study the swarming food chain below — YouTube, indie films, TV serials and insect-scale lip-sync mashups — and not just the tiny apex of tigers. The bottom is where the action is, and where screen literacy originates.

An emerging set of cheap tools is now making it easy to create digital video. There were more than 10 billion views of video on YouTube in September. The most popular videos were watched as many times as any blockbuster movie. Many are mashups of existing video material. Most vernacular video makers start with the tools of Movie Maker or iMovie, or with Web-based video editing software like Jumpcut. They take soundtracks found online, or recorded in their bedrooms, cut and reorder scenes, enter text and then layer in a new story or novel point of view. Remixing commercials is rampant. A typical creation might artfully combine the audio of a Budweiser “Wassup” commercial with visuals from “The Simpsons” (or the Teletubbies or “Lord of the Rings”). Recutting movie trailers allows unknown auteurs to turn a comedy into a horror flick, or vice versa.

Rewriting video can even become a kind of collective sport. Hundreds of thousands of passionate anime fans around the world (meeting online, of course) remix Japanese animated cartoons. They clip the cartoons into tiny pieces, some only a few frames long, then rearrange them with video editing software and give them new soundtracks and music, often with English dialogue. This probably involves far more work than was required to edit the original cartoon but far less work than editing a clip a decade ago. The new videos, called Anime Music Videos, tell completely new stories. The real achievement in this subculture is to win the Iron Editor challenge. Just as in the TV cookoff contest “Iron Chef,” the Iron Editor must remix videos in real time in front of an audience while competing with other editors to demonstrate superior visual literacy. The best editors can remix video as fast as you might type.

In fact, the habits of the mashup are borrowed from textual literacy. You cut and paste words on a page. You quote verbatim from an expert. You paraphrase a lovely expression. You add a layer of detail found elsewhere. You borrow the structure from one work to use as your own. You move frames around as if they were phrases.

Digital technology gives the professional a new language as well. An image stored on a memory disc instead of celluloid film has a plasticity that allows it to be manipulated as if the picture were words rather than a

photo. Hollywood mavericks like [George Lucas](#) have embraced digital technology and pioneered a more fluent way of filmmaking. In his “Star Wars” films, Lucas devised a method of moviemaking that has more in common with the way books and paintings are made than with traditional cinematography.

In classic cinematography, a film is planned out in scenes; the scenes are filmed (usually more than once); and from a surfeit of these captured scenes, a movie is assembled. Sometimes a director must go back for “pickup” shots if the final story cannot be told with the available film. With the new screen fluency enabled by digital technology, however, a movie scene is something more flexible: it is like a writer’s paragraph, constantly being revised. Scenes are not captured (as in a photo) but built up incrementally. Layers of visual and audio refinement are added over a crude outline of the motion, the mix constantly in flux, always changeable. George Lucas’s last “Star Wars” movie was layered up in this writerly way. He took the action “Jedis clashing swords — no background” and laid it over a synthetic scene of a bustling marketplace, itself blended from many tiny visual parts. Light sabers and other effects were digitally painted in later, layer by layer. In this way, convincing rain, fire and clouds can be added in additional layers with nearly the same kind of freedom with which Lucas might add “it was a dark and stormy night” while writing the script. Not a single frame of the final movie was left untouched by manipulation. In essence, a digital film is written pixel by pixel.

The recent live-action feature movie “Speed Racer,” while not a box-office hit, took this style of filmmaking even further. The spectacle of an alternative suburbia was created by borrowing from a database of existing visual items and assembling them into background, midground and foreground. Pink flowers came from one photo source, a bicycle from another archive, a generic house roof from yet another. Computers do the hard work of keeping these pieces, no matter how tiny and partial they are, in correct perspective and alignment, even as they move. The result is a film assembled from a million individual existing images. In most films, these pieces are handmade, but increasingly, as in “Speed Racer,” they can be found elsewhere.

In the great hive-mind of image creation, something similar is already happening with still photographs. Every minute, thousands of photographers are uploading their latest photos on the Web site Flickr. The more than three billion photos posted to the site so far cover any subject you can imagine; I have not yet been able to stump the site with a request. Flickr offers more than 200,000 images of the Golden Gate Bridge alone. Every conceivable angle, lighting condition and point of view of the Golden Gate Bridge has been photographed and posted. If you want to use an image of the bridge in your video or movie, there is really no reason to take a new picture of this bridge. It’s been done. All you need is a really easy way to find it.

Similar advances have taken place with 3D models. On [Google SketchUp’s 3D Warehouse](#), you can find insanely detailed three-dimensional virtual models of most major building structures of the world. Need a street in San Francisco? Here’s a filmable virtual set. With powerful search and specification tools, high-resolution clips of any bridge in the world can be circulated into the common visual dictionary for reuse. Out of these ready-made “words,” a film can be assembled, mashed up from readily available parts. The rich databases of component images form a new grammar for moving images.

After all, this is how authors work. We dip into a finite set of established words, called a dictionary, and reassemble these found words into articles, novels and poems that no one has ever seen before. The joy is

recombining them. Indeed it is a rare author who is forced to invent new words. Even the greatest writers do their magic primarily by rearranging formerly used, commonly shared ones. What we do now with words, we'll soon do with images.

For directors who speak this new cinematographic language, even the most photo-realistic scenes are tweaked, remade and written over frame by frame. Filmmaking is thus liberated from the stranglehold of photography. Gone is the frustrating method of trying to capture reality with one or two takes of expensive film and then creating your fantasy from whatever you get. Here reality, or fantasy, is built up one pixel at a time as an author would build a novel one word at a time. Photography champions the world as it is, whereas this new screen mode, like writing and painting, is engineered to explore the world as it might be.

But merely producing movies with ease is not enough for screen fluency, just as producing books with ease on Gutenberg's press did not fully unleash text. Literacy also required a long list of innovations and techniques that permit ordinary readers and writers to manipulate text in ways that make it useful. For instance, quotation symbols make it simple to indicate where one has borrowed text from another writer. Once you have a large document, you need a table of contents to find your way through it. That requires page numbers. Somebody invented them (in the 13th century). Longer texts require an alphabetic index, devised by the Greeks and later developed for libraries of books. Footnotes, invented in about the 12th century, allow tangential information to be displayed outside the linear argument of the main text. And bibliographic citations (invented in the mid-1500s) enable scholars and skeptics to systematically consult sources. These days, of course, we have hyperlinks, which connect one piece of text to another, and tags, which categorize a selected word or phrase for later sorting.

All these inventions (and more) permit any literate person to cut and paste ideas, annotate them with her own thoughts, link them to related ideas, search through vast libraries of work, browse subjects quickly, resequence texts, refind material, quote experts and sample bits of beloved artists. These tools, more than just reading, are the foundations of literacy.

If text literacy meant being able to parse and manipulate texts, then the new screen fluency means being able to parse and manipulate moving images with the same ease. But so far, these "reader" tools of visuality have not made their way to the masses. For example, if I wanted to visually compare the recent spate of bank failures with similar events by referring you to the bank run in the classic movie "It's a Wonderful Life," there is no easy way to point to that scene with precision. (Which of several sequences did I mean, and which part of them?) I can do what I just did and mention the movie title. But even online I cannot link from this sentence to those "passages" in an online movie. We don't have the equivalent of a hyperlink for film yet. With true screen fluency, I'd be able to cite specific frames of a film, or specific items in a frame. Perhaps I am a historian interested in oriental dress, and I want to refer to a fez worn by someone in the movie "Casablanca." I should be able to refer to the fez itself (and not the head it is on) by linking to its image as it "moves" across many frames, just as I can easily link to a printed reference of the fez in text. Or even better, I'd like to annotate the fez in the film with other film clips of fezzes as references.

With full-blown visuality, I should be able to annotate any object, frame or scene in a motion picture with any other object, frame or motion-picture clip. I should be able to search the visual index of a film, or peruse a visual table of contents, or scan a visual abstract of its full length. But how do you do all these things? How

can we browse a film the way we browse a book?

It took several hundred years for the consumer tools of text literacy to crystallize after the invention of printing, but the first visual-literacy tools are already emerging in research labs and on the margins of digital culture. Take, for example, the problem of browsing a feature-length movie. One way to scan a movie would be to super-fast-forward through the two hours in a few minutes. Another way would be to digest it into an abbreviated version in the way a theatrical-movie trailer might. Both these methods can compress the time from hours to minutes. But is there a way to reduce the contents of a movie into imagery that could be grasped quickly, as we might see in a table of contents for a book?

Academic research has produced a few interesting prototypes of video summaries but nothing that works for entire movies. Some popular Web sites with huge selections of movies (like porn sites) have devised a way for users to scan through the content of full movies quickly in a few seconds. When a user clicks the title frame of a movie, the window skips from one key frame to the next, making a rapid slide show, like a flip book of the movie. The abbreviated slide show visually summarizes a few-hour film in a few seconds. Expert software can be used to identify the key frames in a film in order to maximize the effectiveness of the summary.

The holy grail of visuality is to search the library of all movies the way Google can search the Web. Everyone is waiting for a tool that would allow them to type key terms, say “bicycle + dog,” which would retrieve scenes in any film featuring a dog and a bicycle. In an instant you could locate the moment in “The Wizard of Oz” when the witchy Miss Gulch rides off with Toto. Google can instantly pinpoint desirable documents out of billions on the Web because computers can read text, but computers are only starting to learn how to read images.

It is a formidable task, but in the past decade computers have gotten much better at recognizing objects in a picture than most people realize. Researchers have started training computers to recognize a human face. Specialized software can rapidly inspect a photograph’s pixels searching for the signature of a face: circular eyeballs within a larger oval, shadows that verify it is spherical. Once an algorithm has identified a face, the computer could do many things with this knowledge: search for the same face elsewhere, find similar-looking faces or substitute a happier version.

Of course, the world is more than faces; it is full of a million other things that we’d like to have in our screen vocabulary. Currently, the smartest object-recognition software can detect and categorize a few dozen common visual forms. It can search through Flickr photos and highlight the images that contain a dog, a cat, a bicycle, a bottle, an airplane, etc. It can distinguish between a chair and sofa, and it doesn’t identify a bus as a car. But each additional new object to be recognized means the software has to be trained with hundreds of samples of that image. Still, at current rates of improvement, a rudimentary visual search for images is probably only a few years away.

What can be done for one image can also be done for moving images. Viewdle is an experimental Web site that can automatically identify select celebrity faces in video. Hollywood postproduction companies routinely “read” sequences of frames, then “rewrite” their content. Their custom software permits human operators to eradicate wires, backgrounds, unwanted people and even parts of objects as these bits move in

time simply by identifying in the first frame the targets to be removed and then letting the machine smartly replicate the operation across many frames.

The collective intelligence of humans can also be used to make a film more accessible. Avid fans dissect popular movies scene by scene. With maniacal attention to detail, movie enthusiasts will extract bits of dialogue, catalog breaks in continuity, tag appearances of actors and track a thousand other traits. To date most fan responses appear in text form, on sites like the Internet Movie Database. But increasingly fans respond to video with video. The Web site Seismic encourages “video conversations” by enabling users to reply to one video clip with their own video clip. The site organizes the sprawling threads of these visual chats so that they can be read like a paragraph of dialogue.

The sheer number of user-created videos demands screen fluency. The most popular viral videos on the Web can reach millions of downloads. Success garners parodies, mashups or rebuttals — all in video form as well. Some of these offspring videos will earn hundreds of thousands of downloads themselves. And the best parodies spawn more parodies. One site, TimeTube, offers a genealogical view of the most popular videos and their descendants. You can browse a time line of all the videos that refer to an original video on a scale that measures both time and popularity. TimeTube is the visual equivalent of a citation index; instead of tracking which scholarly papers cite other papers, it tracks which videos cite other videos. All of these small innovations enable a literacy of the screen.

As moving images become easier to create, easier to store, easier to annotate and easier to combine into complex narratives, they also become easier to be remanipulated by the audience. This gives images a liquidity similar to words. Fluid images made up of bits flow rapidly onto new screens and can be put to almost any use. Flexible images migrate into new media and seep into the old. Like alphabetic bits, they can be squeezed into links or stretched to fit search engines, indexes and databases. They invite the same satisfying participation in both creation and consumption that the world of text does.

We are people of the screen now. Last year, digital-display manufacturers cranked out four billion new screens, and they expect to produce billions more in the coming years. That’s one new screen each year for every human on earth. With the advent of electronic ink, we will start putting watchable screens on any flat surface. The tools for screen fluency will be built directly into these ubiquitous screens.

With our fingers we will drag objects out of films and cast them in our own movies. A click of our phone camera will capture a landscape, then display its history, which we can use to annotate the image. Text, sound, motion will continue to merge into a single intermedia as they flow through the always-on network. With the assistance of screen fluency tools we might even be able to summon up realistic fantasies spontaneously. Standing before a screen, we could create the visual image of a turquoise rose, glistening with dew, poised in a trim ruby vase, as fast as we could write these words. If we were truly screen literate, maybe even faster. And that is just the opening scene.

Kevin Kelly is senior maverick at Wired and the author of “Out of Control” and a coming book on what technology wants.

[Privacy Policy](#) | [Search](#) | [Corrections](#) | [RSS](#) | [First Look](#) | [Help](#) | [Contact Us](#) | [Work for Us](#) | [Site Map](#)

What Technology Wants: - kevin-kelly-part-0/ From The New York Times Idea Lab November 23, 2008 Becoming Screen Literate By KEVIN KELLY Everywhere we look, we see screens. The other day I watched clips from a movie as I pumped gas into my car. We are becoming people of the screen. The fluid and fleeting symbols on a screen pull us away from the classical notions of monumental authors and authority. On the screen, the subjective again trumps the objective.