

19 April, 2015

Dear cixa,

You asked what else I've been up to, and although this mail is addressed to you, I can already see that I should post it online in case anyone else is wondering. But maybe I should start with *what*, rather than *what else*, I've been up to. Although my work can seem hopelessly fragmented and scattered, in fact I consider myself a very poor multitasker. Everything I do is part of the same project, is an attempt to express & externalize a working set of ideas and to move in a single direction. The PDF I sent you for your collection, [training.pdf](#), is a 10,000-item training set I constructed by perspective-warping pages from [MuckRock's](#) FOIL archive:



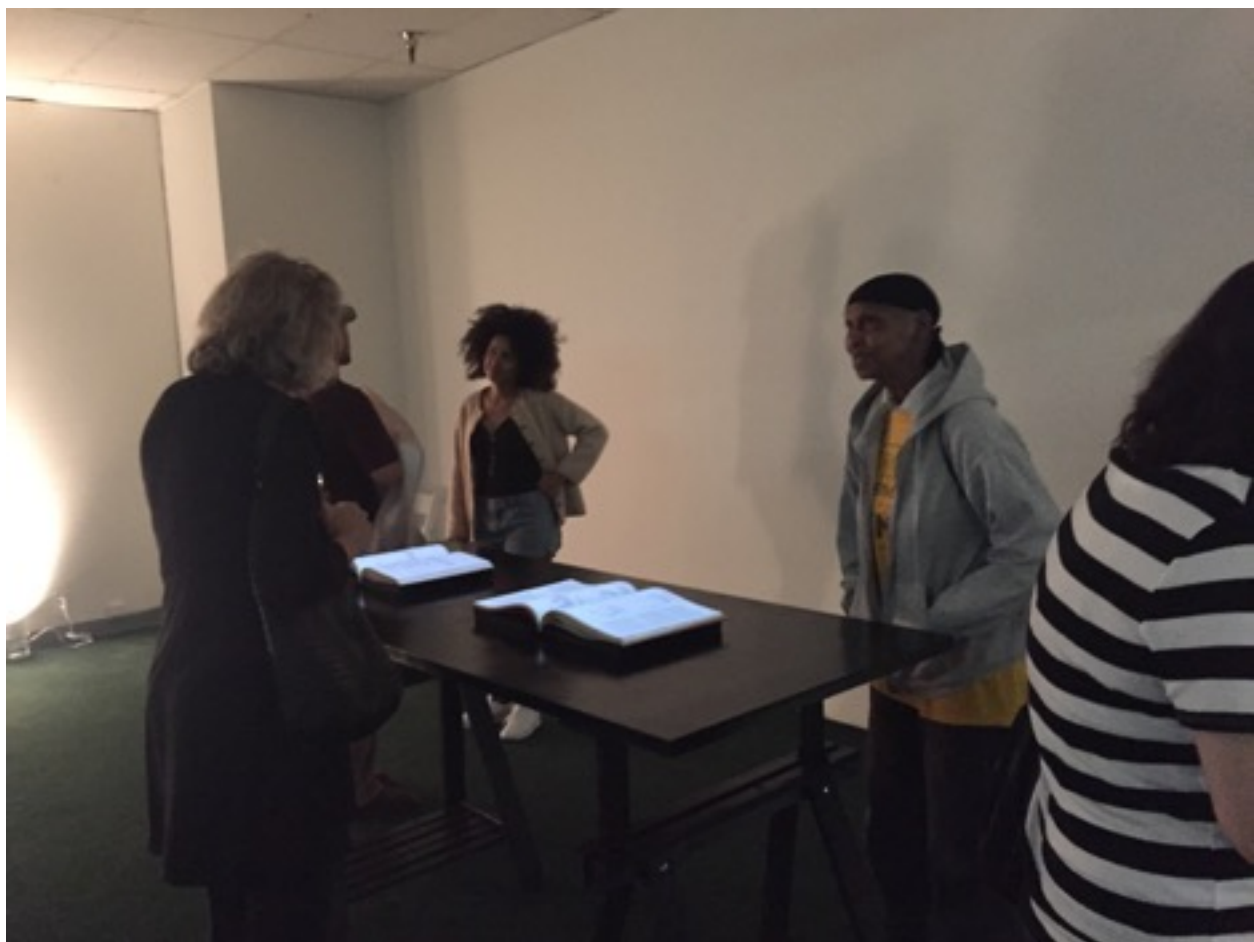
I would like to train an AI to detect the orientation and corners of these pages. Unfortunately, I am a dilettante when it comes to modern AI and computer vision, so I don't know if I will succeed in this particular task, at least with this approach, but I was blown away by a [journal article](#) that I read about "natural" scene text recognition. Instead of the old-school computer vision approach of writing hundreds of thousands of lines of optimized C++ to specify a procedure for turning pixels into letters, or instead of the new big-data-deep-learning approach of feeding millions of labeled datapoints into a neural net, Jaderberg *et al* have twisted the problem into *simulation* of arbitrary amounts of realistic data. None of these approaches give us profound/legible insight into how visual systems (could) work, but the latter at least proposes to

tell us how the world has come to be as it is. And that seems like an interesting thing to try and specify.

I am re-reading Baudrillard's *Simulations* and finding uncanny resonance.

Ultimately, I am not interested in designing The Matrix. On the contrary, my foray into computer vision techniques is motivated by the opposite impulse. Under the influence of [Bret Victor](#) and others at my [research lab](#), I have had to face a longstanding and unquestioned assumption in my work that working with digital media implies reception through a (usually networked) computer on a (usually tiny) screen with mouse/keyboard/(sometimes touch) inputs. I have been thinking about the physical and the virtual, and how to retain some of what is good about the former without denying the endless possibilities of the latter. Is it trite to seek out "reality"?

Last Saturday—this is my excuse for being so late with your PDF—was the opening for a show I worked on in collaboration with the [Los Angeles Poverty Department \(LAPD\)](#). At the first exhibition to inaugurate the new Skid Row History Museum in downtown LA, we presented the *Blue Book* and the *Silver Book*, dueling urban planning documents from the 70s, as entry into a growing (digital) archive about Skid Row. But look! These are no PDFs on the table:



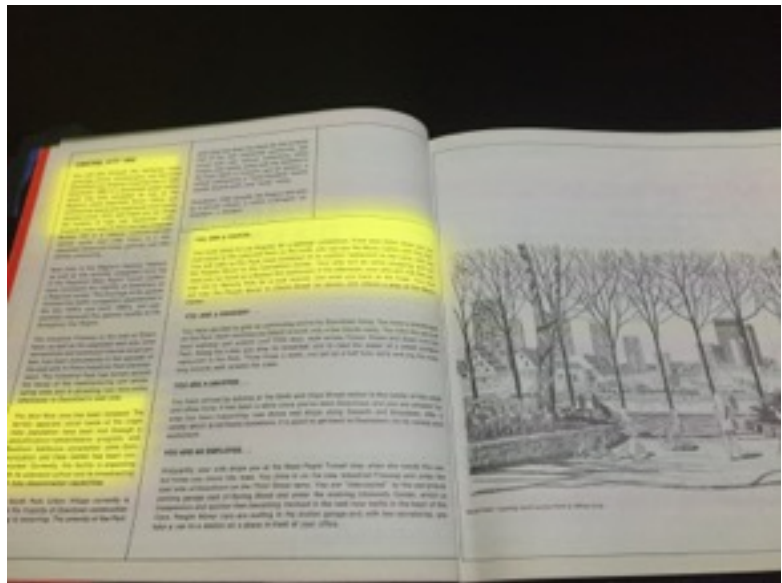
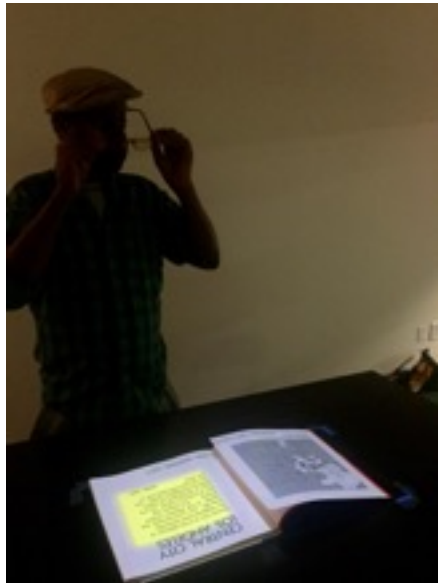
They are books. (Original, hard-to-find books, in fact.) You can stand around them, see them from behind, touch them, leave your finger in one page while you look at the other, &c.



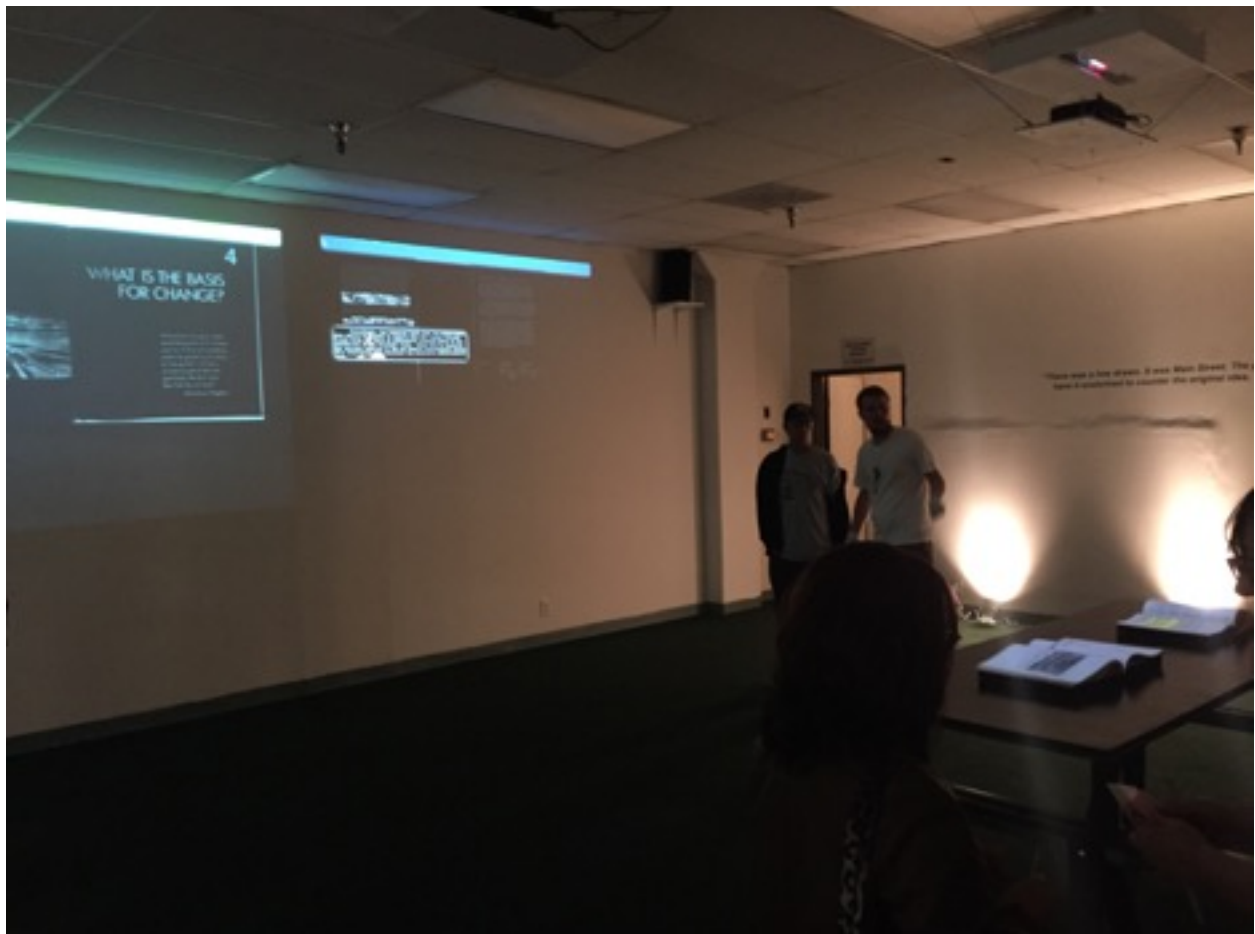
But they are not *just* books. If you look at the top of this photo, you can see a strange glowing box above the books. This is a device I designed (with some help!) containing a mirrored projector and a HD camera:

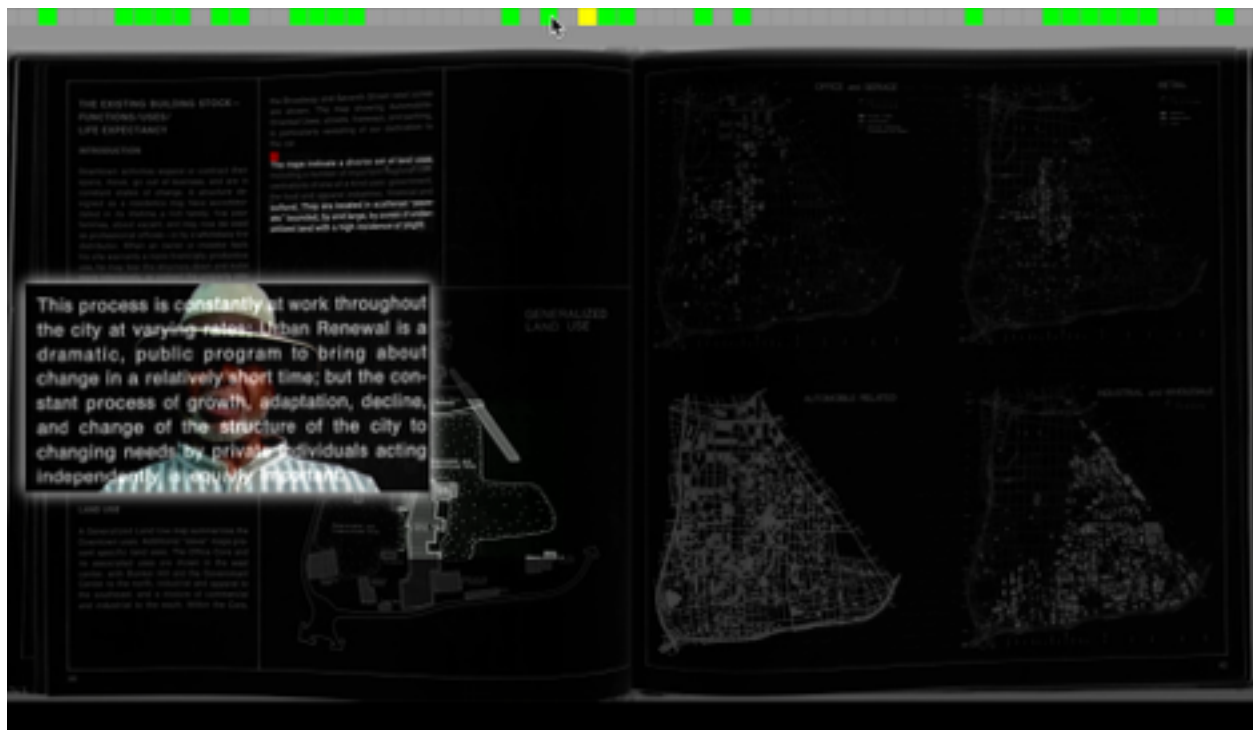


With the camera, I can detect the currently-open pages, and with the projector I can augment the books with page-specific highlights:



Then, in addition to highlighting the books, pages may trigger video projections in the room with LAPDers reading (sometimes egregious) paragraphs from the books or external archive material and interviews providing context/counter-narrative.





I made a [short clickthrough](#) of what the projection looked like, though of course it's a completely different experience within the space (whether you are behind the book or not). The book-highlight-projection format isn't really worked out yet, but it's a start.

It was a struggle getting the page-detection working, but it ended up surprisingly robust. Here's a [video](#) of me flipping through the *Silver Book*. Although I was able to get away with some naïve approaches in prior studies (for this [video binder](#) that switches video based on the current page, taking the average color of each thumbnail was good enough to sort-of work), this required more research and high-end techniques. Inspired by [Tomasz Malisiewicz's Exemplar SVM research](#), I preprocessed a photo of each spread into a thresholded 50x25 image and built a classifier that separates each spread from the rest. Here's some of the training data for the Silver Book (note the hands hovering over many of the images—occlusion was not an issue at all):





Despite some CV success, I didn't do any automatic page alignment so we had to design a proper jail for the books:

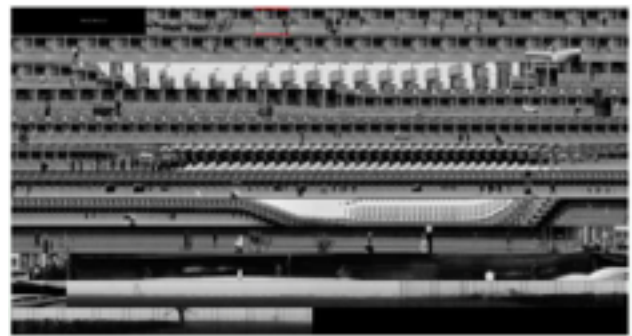
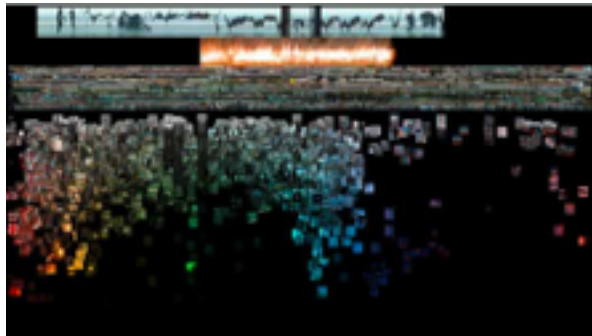


Making the physical world conform to the needs of the virtual can be brutal. That's LAPD's Austin Hines, left, holding open a binder clip that's screwed into the table to receive and fix the *Silver Book*. I hope the page-warp simulations I sent you can ultimately lead to a gentler interface between physical and virtual. I had also hoped that we could *construct* the links out of the books with a physical interface, but I couldn't get it together in time (we didn't even *find* the right books until days before the opening) so I fell back on throwing together an ad-hoc web app:



One interesting idea that slipped into this UI are the video columns, left bottom. With the drop-caps, it's an explicit typographic flair in my longstanding attempt to make video behave a little more like text (*e.g.*, video should be skimmable, fixed & stable w/r/t the passage of time, and cut-copy-paste tangible/mutable). You can see the interaction [here](#), with all its present clumsiness. The provenance of this half-baked UI element is certainly more interesting than its current form, and I will try to unpack that history, if you will indulge some backstory. (Despite a recent [stalled effort](#), I still lack proper documentation for my video prototypes, so this will need to cover some old territory.)

From the first slitscan renderings I made several years ago—accidentally, as I kept cropping video more and more, down to a single pixel-column—I have been obsessed with the deep confluences of time and space. In these old screens, the slitscan triumphs over all the other ideas.



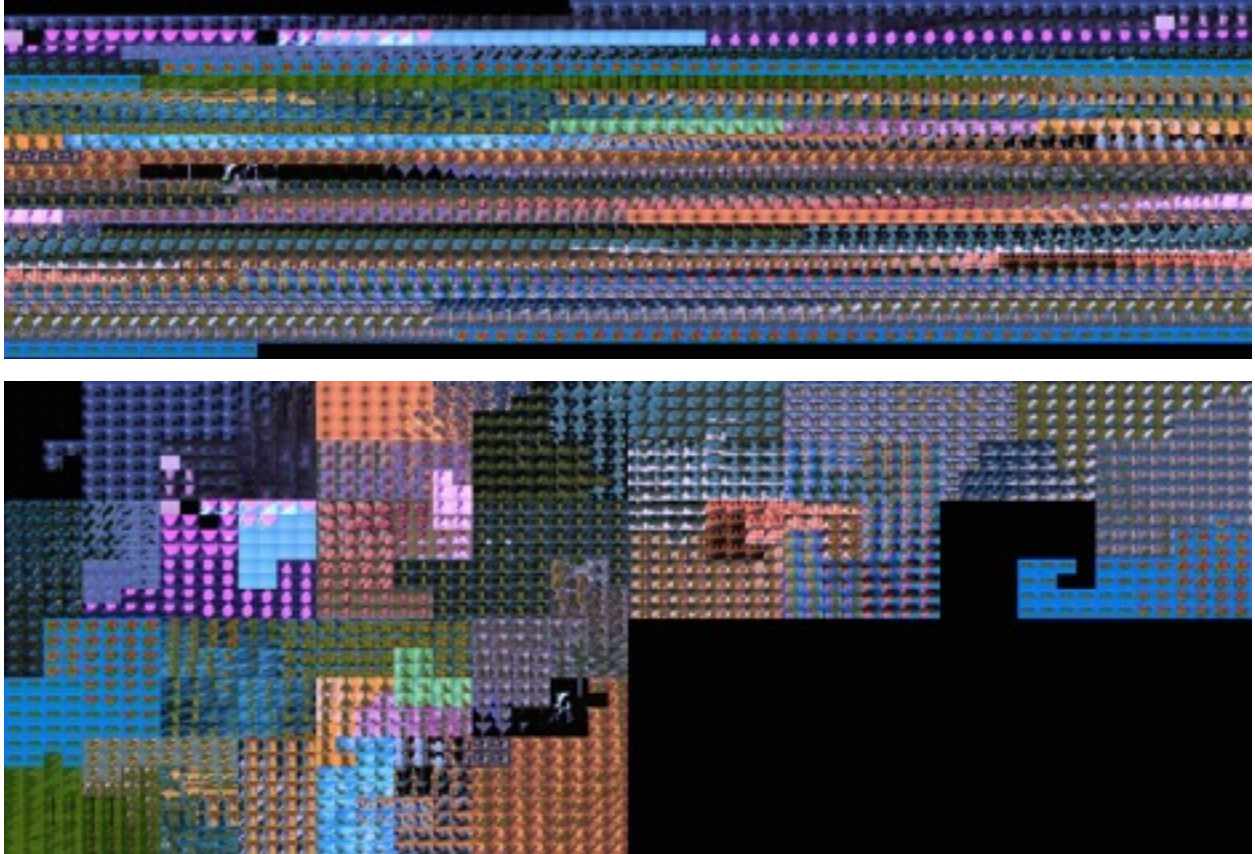
In the [GOD prototype](#) (2012, left), with 15 Godard films visualized one-per-line and each cut scattered by hue/saturation, it is the frozen choreography of the dancer's path as she slides through her fixed routine that seemed to hold the most promise for a future cinema. And in [Chewing](#) (2013, right), it was the flattening of the present into a weirdly expressive spatial timeline that opened up John Smith's 1976 film *The Girl Chewing Gum* to a new form of cinematic authority.

Slitscans may be good for some things, but sadly they did not turn out to be a panacea for temporal visibility (they don't zoom well, and they rely on certain very specific forms of motion/stillness to be effective). So I moved on to other representations. The [GPU timelines](#) I've been working on seem very powerful, and towards the end of last year, I put together a [study](#) very explicitly linking a video playhead to a floorplan location (on a fundraising video shot with a moving camera).

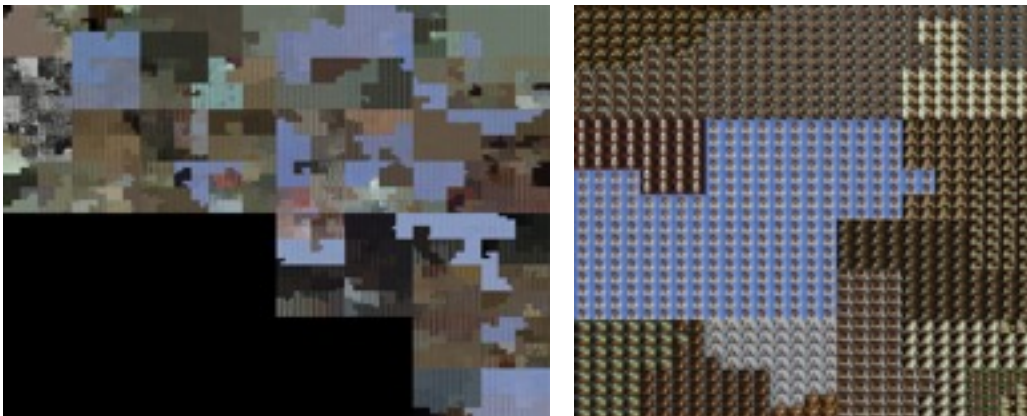




When I showed this to my friend [Bryan Newbold](#), he pointed me to Hilbert curves as a formal, mathematical approach for translating one dimension (a line, *e.g.*, a timeline) into two or more dimensions (a space). Thinking back to [Dave Cerf](#)'s obsession with video grids, where each frame of a video is arranged one after another, line-breaking to form a grid, I saw promise in Hilbert curves to address the naïve grids' biggest deficiency, the creation of arbitrary patterns as a result of line breaks. Here is a cinematic classic rendered first as a left-right grid, and then as a Hilbert grid (and a gif—another classic file format—moving between the two):

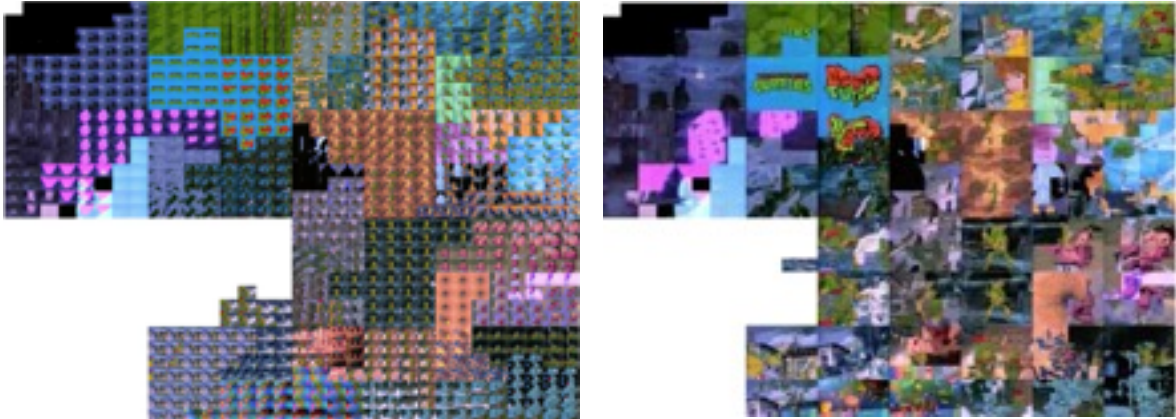


I loved how, in the Hilbert arrangement, cuts turned into “regions” of a videogame-like landscape. This property holds at a wide range of timescales. Here are 45,000 frames in John Berger’s *Ways of Seeing* (ep.1), left, with a detail, right:



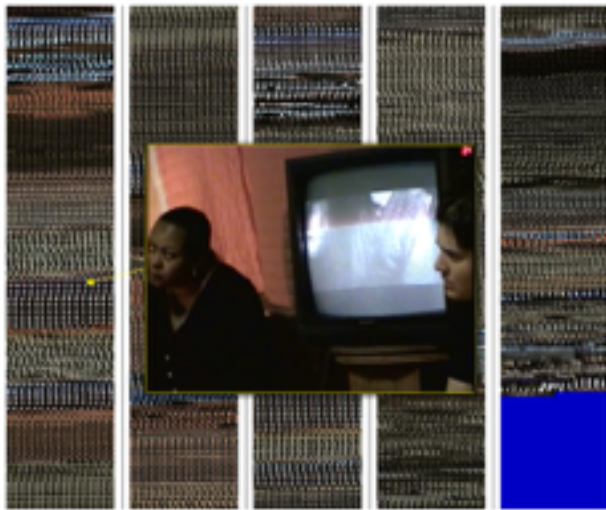


I pushed the Hilbert grids a little further, making a psychedelic fractal zoom:



and animating the whole grid, but I think it was Chaim Gingold<sup>1</sup> who figured out that most of the benefits of the Hilbert grid could be achieved with newspaper-style columns. So for the next version of InterLace, in development, I am trying to use columns for video, PDFs, and text.

153\_2000\_Red Beard-Red Beard\_Courtland H\_Perf\_VH8



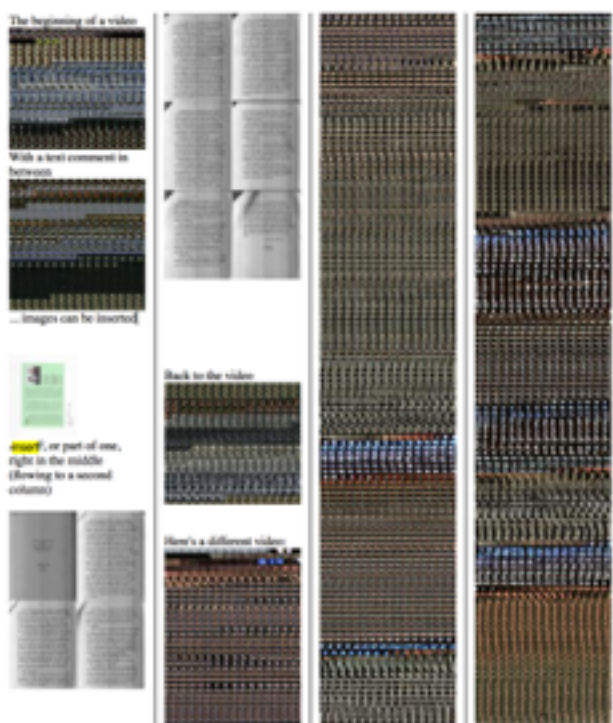
5011002815\_1501261609600\_ocr



There's still an absolutely daunting amount of work to do before anything usable will come of this, but *something* is starting to happen:



<sup>1</sup> Losing track of the exact source of particular ideas is a new feeling, and is testament to the depth of conversation I enjoy with my colleagues at CDG. I am spoiled.



Probably these screen-based UI's are more compelling in this format—flattened on a page, glowing on a screen, &c—but even in the early stages of my physical explorations I have caught glimpses of something that seems far more powerful (and humane) than anything possible in a tiny rectangle. How to describe the sense of balance that (literally) standing in for a page of a book or the speed of a video provides? Do any of these words express the comfort and confidence that replacing one's filesystem with physical links in space can give?

As ever, my work and life are hopelessly blurred, and my attempts to work in physical rather than purely virtual space have been rocky. I tried to move into a converted shipping container (left), as practice modifying my physical environment, but zoning mishaps ended that experiment. Likewise with car custodianship (right) and engine mishaps. I look back at how I used to live—drifting with my life in a backpack wiring into my laptop on the trans-Siberian rail—and in retrospect see how *easy* (“frictionless”) that was. The physical world is a challenge, requiring constant maintenance, but I think it may be worth the effort.



Your correspondent,  
R.M.O.