Amy Alexander

SVEN (Surveillance Video Entertainment Network): Looking Back and Forward

2017

https://doi.org/10.25969/mediarep/4034

Veröffentlichungsversion / published version
Sammelbandbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:
Alexander, Amy: SVEN (Surveillance Video Entertainment Network): Looking Back and Forward. In: Annette Brauerhoch, Norbert Otto Eke, Renate Wieser u.a. (Hg.): *Entautomatisierung.* Paderborn: Fink 2017 (Schriftenreihe des Graduiertenkollegs "Automatismen"), S. 67–75. DOI: https://doi.org/10.25969/mediarep/4034.

Erstmalig hier erschienen / Initial publication here:
https://nbn-resolving.org/urn:nbn:de:hbz:466:2-28545

AMY ALEXANDER

SVEN (SURVEILLANCE VIDEO ENTERTAINMENT NETWORK): LOOKING BACK AND FORWARD

In 2006, with collaborators Wojciech Kosma and Vincent Rabaud, I launched a project called SVEN (Surveillance Video Entertainment Network). SVEN was originally a van-based street performance, and we eventually developed an installation version for museums and galleries. The premise of SVEN was, "If computer vision technology can be used to detect when you look like a 'terrorist', 'criminal', or other 'undesirable' – why not when you look like a 'rock star'?"

1 – SVEN on the streets of Zurich. Guest of Digital Arts Weeks, 2006

We developed SVEN to behave most of the time like a normal CCTV system – scanning the area and displaying surveillance video on three monitors – a large main screen with two adjacent smaller screens. This setup emulated CCTV monitoring consoles, which often have a large screen for primary monitoring accompanied by adjacent smaller screens for monitoring remote areas in which miscreants are likely to appear. However, SVEN's computer vision software wasn't programmed to look for people who appeared to be terrorists or shoplifters – it was programmed to look for people who appeared to be "rock stars". When SVEN found someone, who, in its algorithmic opinion, resembled a rock star, the SVEN software would process the live surveillance video on the main screen to look like a music video. The smaller screens

would then display a freeze frame "mug shot" image of the target person and, for comparison, a still image of the rock star whom they had "matched."

2 – SVEN production still

Why surveillance?

Surveillance video was a big concern in 2006. Five years after the attacks of September 11, 2001, the public was arguably at the height of distrust of government officials. In the US, the public's initial acquiescence to increased security immediately after the attacks had given way to resentment and mistrust in the wake of repeated reports of racial profiling and bureaucratic snafus that targeted innocent people rather than actual terrorists. Many people feared the government would use whatever means they had at their disposal to target people by ethnicity, political leanings, or any other reason that suited them. Technology, meanwhile, was and is often seen as some sort of mysterious, magical power, despite its having pervaded the everyday on many different levels.

Engineers refer to systems whose inner workings cannot be observed as "black boxes". You can know what goes into and out of them, but only the designers know what happens in between. The user doesn't know how the system works and so is disempowered. This disempowerment through concealment is in some ways similar to the function of the Panopticon. As originally described by eighteenth century philosopher Jeremy Bentham, a Panopticon is a prison designed to let authorities observe the inmates without their knowing whether or when they are being watched. In the twentieth century Michel Fou-

cault expanded Bentham's model to apply to hierarchical institutions in general: just like Bentham's prison inmates, workers and schoolchildren submit to the power of unseen authority. They expect they are being watched at some times – but when? In recent years, as the use of surveillance video has become ubiquitous; its use has been referred to as panopticism. But again, the focus is on the question of time. If a human operator is watching the monitor, we know how we are being observed by authorities – we just don't know when. Since we cannot know when we are being watched, our behavior is influenced at all times by unseen power.

But I'd argue that with technology-based observation – like computer vision-based surveillance – *how* we're being observed is just as important as when. Unlike humans, who must generally be paid for time spent observing, software can run all the time. But how will it decide who among us is "misbehaving"? Paranoia abounds as we imagine computer vision software programmed to detect skin color, facial hair, bohemian dress. But in reality, we rarely know how these systems work; they're presented to us as black boxes. The air of mystery helps perpetuate a climate of fear around identity recognition technology. When you're in fear, you can't move. And the technology becomes the sole domain of people who would use it to scare those they want to control – whether they actually have the ability to do scary things with it or not.

Defamiliarizing the Black Box

We may not always be able to see inside black boxes. But when we recognize them as such – when we look straight at them instead of past them – they lose some of their "mysterious" powers. My idea with SVEN was to separate the technology of computer vision-based recognition from the politics – the "what" from the "how". It's not the technology itself that's frightening; it's the fact that we envision it being used against us in frightening ways. But we could use it ourselves in ways that would be more enjoyable – even amusing. Instead of detecting "undesirable" people we could use it to detect "desirable" ones. And who could be more desirable than rock stars?

Of course, pop culture constantly inundates us with the assumption that rock stars are among the most desirable of humans. But there's another reason SVEN converts surveillance videos to music videos: the two seem to be different sides of the same coin. Music videos are often shot cinéma vérité style, giving the illusion that they depict their stars in "real life" situations – from walking the streets to hanging out backstage before a concert. The stars appear to be unaware of the camera, and their seeming to be under the camera's surveillance becomes part of the video's "cool" aesthetic. On the other hand,

there have long been people who performed for surveillance cameras – from the deliberate public performances of the Surveillance Camera Players[1] to the impromptu performances of millions of people who notice a surveillance camera and smile, wave, or even dance for it. There's a fine line between surveillance and exhibition – and a fine line between surveillance and music video.

How can we tease out the "how?"

Computer vision research often focuses on algorithmically spotting undesirable behaviors – is this person stealing[2] or behaving aggressively?[3] – and identifying undesirable people – is this person's image in a database of terrorists?[4] But detecting potential rock stars algorithmically has something in common with detecting potential terrorists algorithmically. Humans can innately differentiate one person from another (most of the time). But this very basic human capability has to be somehow described quantitatively to a computer. Is such a task even possible? If so, what criteria do we use? These decisions have to be made by humans, and they have ethical consequences. Although it's convenient to think of software and algorithms as neutral and mechanical, creating software is a creative human endeavor, and consequently, algorithms are subjective.

In "Face Recognition Using Eigenfaces"[5] researchers documented their attempt to use computer vision algorithms to match photographs of individuals with those in a database. We see that the algorithm detected the correct person from the database in a large percentage of cases. But the situations in which it erred are interesting. The system could be fooled by lighting, facial hair, glasses or even facial expressions; people of different races could sometimes be incorrectly matched based on similar smug smirks. Could we one day be pulled aside in the airport for being spotted on camera with the wrong attitude? So it seems that not only the programmers' subjectivity is a factor here, but also the software's. Software's tendency to sometimes be subjective – to do what it wants to do – is more commonly known as "bugginess".

---

1 "Surveillance Camera Players", n.d., http://www.notbored.org/the-scp.html, last downloaded 2014-01-01.

2 "StopLift Checkout Vision Systems", n.d., http://www.stoplift.com/products/sweetheart-detection/, last downloaded 2014-01-01.

3 Dejan Arsic/Björn Schuller/Gerhard Rigoll, *Suspicious Behavior Detection in Public Transport by Fusion of Low-Level Video Descriptors*, http://www.mmk.ei.tum.de/publ/pdf/07/07ars1.pdf, last downloaded 2014-01-01.

4 "Candid Camera: Computer-Based Facial Recognition System Spots Terrorists Entering the U.S.", n.d., http://www.odu.edu/ao/instadv/quest/CandidCamera.html, last downloaded 2014-01-01.

5 "Eigenfaces – Christopher de Coro", n.d., http://www.cs.princeton.edu/~cdecoro/eigenfaces/, last downloaded 2014-01-01.

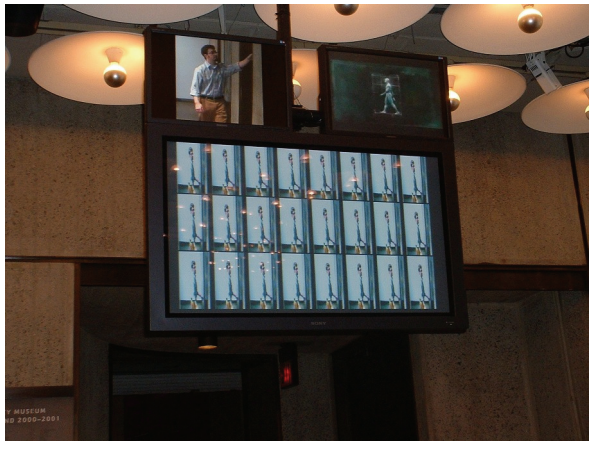
So in developing the algorithms by which SVEN detected rock stars, we focused not on getting it right – there's little chance of SVEN actually catching Bono – but on getting it wrong in ways that highlighted the subjectivity and bugginess inherent in computer vision systems. Computer vision researcher Vincent Rabaud wrote the custom computer vision software, called SVEN CV, which is available for download along with source code.[6] SVEN CV tracks people as they move through the camera's view. If there are multiple people in frame, it keeps track of which person is which: only one person at a time can be the "star". SVEN CV determines the star's position in the frame and the outline of their body, and it segments the star's body to determine the position of their head, torso, and legs. This positional data will be used by SVEN's video processing software, written by media artist Wojciech Kosma,[7] to frame cinematic shots like "close-up", "long shot", etc. and to position visual effects in appropriate places on the screen. SVEN CV then tries to find a facial expression: it does this by comparing the person's face to composite data stored in the program. This stored data was generated by photographing a number of control subjects, who were asked to perform various specific facial expressions that appear frequently in music videos (squinting, pouting, singing, smiling etc.) A composite of all control subjects' data was then created for each facial expression, using a technique similar, though not identical, to the way Eigenfaces works. Of course, software written using Eigenfaces or other facial recognition methods usually tries to ignore variations caused by facial expressions, because its authors have decided expressions hinder the task of spotting "bad guys": humans understand expressions as behaviors, but to a computer, they can appear to be part of a person's identity. In contrast, SVEN CV actually tries to spot expressions, because we decided they were useful to us in detecting "rock stars". Other characteristics SVEN CV looks for include hair color, clothing color, direction of movement (from which we also determine which way the person is facing), and glasses (sunglasses or otherwise). The SVEN video software then compares these results against a database of music video star "mug shots" and determines whether it has found a match. If you've got dark hair, are wearing black, and smirk like Bono in Vertigo, you're as good as guilty.

Are any of the characteristics SVEN tests for actually useful in identifying specific rock stars? Not likely. When viewers see freeze frame images of the real and "spotted" rock stars side by side on SVEN's screens they realize there are certain similarities between the faces, but that only a computer program could conclude that these add up to the same person. The transparent failure of the algorithmic comparisons is what makes it funny. While "real" recognition

[6] "SVEN Computer Vision Software", n.d., http://deprogramming.us/sven/software.html, last downloaded 2014-01-01.

[7] Jesse Gilbert wrote the video code for the initial prototype of SVEN, and Nikhil Rasiwasia wrote the prototype computer vision code. Cristyn Magnus and I wrote additional video processing code for the final production version.

software often uses more sophisticated algorithms, and those algorithms generally aren't designed to do funny things,[8] it still comes down to algorithms. We hope that SVEN can help reveal that algorithms aren't magic: people can question how exactly they're supposed to work and what their limitations are.



3 and 4 – SVEN at the Whitney Museum, New York, 2007

[8] It's worth noting that Vincent Rabaud's work on the SVEN computer vision software was funded by Cal-IT2 (California Institute for Telecommunications and Information Technology) at University of California, San Diego. Cal-IT2 saw the research Vincent was doing in detecting people's "desirable" characteristics as potentially useful to the institute's computer vision research efforts. So although we used SVEN CV to do some funny things, it really isn't a joke.

Sticking it to "The Man?"

When an art project has a political theme, it can be helpful to consider its purpose. Can it be called a weapon? Can it raise awareness of an issue? Help us vent anger and frustration? It is important to understand that SVEN does nothing to stop authorities from misusing surveillance or computer vision technologies. We'd be deluding ourselves dangerously if we thought we'd accomplished that. What it tries to do is help make the technology less scary by using humor to separate it from its scary applications. Once we understand that technology is both accessible and fallible, then we can begin to fight it directly.

Where does that leave us now?

It's been five years since we created SVEN, and some of the issues around surveillance and software have shifted. The notion of surveillance as something that's done *to* us has shifted drastically over the past few years. Video cameras on mobile phones have turned the general public into surveillance camera operators who can potentially record the next viral video at a moment's notice. The manifold documentation of the so-called Arab Spring is a remarkable case in point. Video footage shot during those protests – like that of the shooting death of Neda Agha-Soltan in Iran in 2009 – or in other recent protests – like the video of a US university police officer pepper spraying a row of seated students during the "Occupy" protests in 2011 – has done more than change the dialogue around the protests. It has put authorities at the center of a kind of inverted panopticon. But in this case, the power doesn't derive from the targets not knowing when they are being watched – with this many sets of eyes, video cameras and YouTube accounts, they almost always are. Even "the man" can't hide now.

So perhaps we don't need to worry about *how* we're being watched after all?

It would be nice if simply putting video cameras into the hands of the public were all we needed to turn the surveillance tables on the authorities. But our problems with misuse of computer vision recognition technology may be just beginning. For example, in the US state of Massachusetts, a facial recognition system designed to scan driver's license photos for terrorists and fraudsters misidentifies a thousand innocent people a year. Many of these people have their licenses suspended; all have to endure the bureaucracy of proving their identity. Officials have demonstrated a cavalier attitude toward the dysfunc-

tionality of their system, emphasizing that it protects the public by "securing the identity of 4 ½ million drivers", and dismissing erroneous matches as "inconveniences" to those misidentified.

But who – or what – is making those mistakes? And how? What criteria does the system use to decide on a face match? Does it get fooled by lighting? Beards? Smirks and pouts? Do officials ensure that humans oversee that the system is being used properly? Or double check its results? When human law enforcement officers stop and frisk innocent people, we know enough to at least ask questions about what criteria they use. Profiling by humans is clearly an ethical issue. But when a computer does it, it's often dismissed as a mistake or bug. Computers have no ethical responsibility – but the people who program them, buy them and operate them do.

So while surveillance video has escaped the proverbial black box of unexamined acquiescence to Big Brother, identity recognition technology remains for now, inside it. What brought surveillance video into the wider consciousness was the public's access to it. When, through consumer video devices, people became surveillance camera operators themselves, they learned about how the process worked. Will the same happen with computer vision recognition? Face recognition functionality now appears in everything from photo applications to Facebook. We may not all have the technical knowledge to produce face recognition algorithms as we can with surveillance video, but we can start to understand the ways in which it works – and doesn't work. Once you've had the experience of Facebook identifying a photo of a cartoon character as your brother, you realize there are some questions to be asked.

Will a "read-only" understanding of a technology be enough? Can we be literate in any medium without knowing how to produce it? Or will everyone need to learn to program before we can move toward a technology-literate society? Looks like we're about to find out.

See you back in another five years.

Literature

Arsic, Dejan/Schuller, Björn/Rigoll, Gerhard, *Suspicious Behavior Detection in Public Transport by Fusion of Low-Level Video descriptors*, http://www.mmk.ei.tum.de/publ/pdf/07/07ars1.pdf, last downloaded 2014-01-01.

"Candid Camera: Computer-Based Facial Recognition System Spots Terrorists Entering the U.S.", n.d., http://www.odu.edu/ao/instadv/quest/CandidCamera.html., last downloaded 2014-01-01.

"Eigenfaces – Christopher de Coro", n.d., http://www.cs.princeton.edu/~cdecoro/eigenfaces/, last downloaded 2014-01-01.

"StopLift Checkout Vision Systems", n.d., http://www.stoplift.com/products/sweet heart-detection/, last downloaded 2014-01-01.
"Surveillance Camera Players", n.d., http://www.notbored.org/the-scp.html, last downloaded 2014-01-01.
"SVEN Computer Vision Software", n.d., http://deprogramming.us/sven/software.html, last downloaded 2014-01-01.