

The video codec formerly known as Fermi

AUTHOR: Steven Robertson
CONTACT: steven@strobe.cc
PUBLISHED: 2009-10-30
TAGS: Codec, Article

ABSTRACT

My thesis involves creating a video codec based on a new kind of image transform. In this article, I tell you about how I plan to tell you about the theory of how it works. Not that useful, yeah, but at least I can get some of the invective off my chest before I start diving into technical details in later posts.

I'm writing a video codec!

To those unfamiliar with the video compression landscape, this seems like a bold and innovative move, and one which should generate much excitement from people who have that irrational love of squeezing their treasured collection of videos of their cats doing silly things in high definition down by an additional 2%¹. On the other hand, more experienced individuals might simply mutter, "another one?"

Yes, folks, another stab at video compression, a field which generates hundreds of published papers a year (and likely many more unpublishable ones) describing how to eke out another 0.02dB PSNR from a Playboy centerfold snapshot or a ten-second CIF-sized video of some UT idiot fumbling a football². Hooray, you've saved five bytes on one video and increased decoding time by 600%. *In MatLab*.

Okay, I admit it, that's unfair. Engineering research is all about trying new things, even if they sometimes kinda suck, and publishing those papers may eventually lead to better real-world codecs. I can understand and almost forgive people trying to make tenure by trying ten ideas in a MatLab script and spacing out twenty 4-page papers over two years describing them. It should also be said that video compression is one of those fields in computer science where things get *harder* as time goes on: because compression is fighting against entropy, each step forward takes us another step towards a hard, nature-of-the-universe law, making it that much harder to attain the next round of stunning performance enhancements.

Of course, sometimes a new idea comes along which changes something fundamental about the way we do things and enables that next round of gains. For a couple years, a class of domain transforms which can be accurately but somewhat enigmatically be referred to as "directional multiresolution decompositions" have seemed like they could be that breakthrough in image and video compression. I'm preparing a blog post describing that... *very impressive-sounding* concept in high-level terms³, but for now it will suffice to say that these methods are new, promising, and just different enough to require throwing out most of the old tricks we used to squish video before.

¹I'm one of 'em. (The 2% part, not the cat videos. I do not have cats.)

²Not kidding.

³After describing it to friends and family for a few months now, I might be able to pull this off in a reasonably articulate way.

As far as my preliminary paper-trawl has turned up, nobody's even tried a directional decomposition video codec before, much less constructed one with a real-world implementation that makes impressive gains in coding efficiency. This either means that the topic is ripe for the researching, or that people have tried and failed abjectly. It certainly means that the topic is challenging.

Naturally, this means I pretty much have to try it, because I'm An Idiot™.

So, in one way, this is slightly different from most of the instances in which someone announces that they're making a new video codec before they have code to prove it, as the fundamental underpinnings of the video codec are substantially different from any previous attempts I have seen and thus have the potential for producing significant and informative results.

In another way, it's *exactly* like most of those instances, because by creating a new codec I'm ensuring that it will be at least ten years or so before it sees widespread adoption and general usefulness, *even if it is a staggeringly good improvement*. Perhaps it will take even longer than that. But, alas, there's little room for avoiding that fate, as these changes can't simply be patched into the [Dirac](#) or [x264](#) code.

As I move from researching this problem to actually coding it, I hope to collaborate closely with existing open-source video codecs, -stealing- sharing code wherever possible and submitting patches enabling any new backwards-compatible compression techniques I might be so fortunate as to discover.

More as I learn it.

ABOUT THE TITLE

Oh, and I should mention: the codec was to be called "Fermi". Enrico Fermi derived Fermi-Dirac statistics independently of Paul Dirac, and since I plan to make use of as much code (including a significant amount of the bitstream specification) from Dirac as possible, I thought the name was apt. Except NVIDIA decided to steal my name for their latest GPU architecture. The product is still months away from launch, but Googling for "fermi video" already fills the page with noise about the cards, so a new name is required but not yet chosen⁴.

⁴I'm calling it "The Video Codec Formerly Known As Fermi" in my head. FCAF (eff-kaff) for short.