12M frames per second
How I got into a Netflix movie

Based on a true story :)}
If brute force doesn’t work you aren’t using enough
Where it all began

Background

Employee #1 at Viewdle face recognition in video startup out of Kiev, Ukraine
Acquired by Google/Motorola

Got to LA to pitch face recognition to movie studios.
Met every single studio.
They never needed face recognition.
Face recognition is cool
Can you compare two video files?
### Revisions

<table>
<thead>
<tr>
<th>Revision</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>zhuker created this gist a minute ago.</td>
</tr>
<tr>
<td>19</td>
<td>zhuker revised this gist 24 seconds ago.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>zhuker</td>
<td>edited</td>
<td>diff.txt</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Changes</th>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1,3</td>
<td></td>
<td>@ @This part of the document has stayed the same from version 8 to 10.</td>
</tr>
<tr>
<td>+1</td>
<td></td>
<td>+1 This part of the document has stayed the same from version 8 to 10.</td>
</tr>
<tr>
<td>-8,17</td>
<td></td>
<td>@ @This part of the document has stayed the same from version 8 to 10.</td>
</tr>
<tr>
<td>+14,16</td>
<td></td>
<td>+14,16 This part of the document has stayed the same from version 8 to 10.</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>+1 This is an important notice! It should therefore be located at the beginning of this document!</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>+14 This is an important notice! It should therefore be located at the beginning of this document!</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>+8 compress the size of the changes.</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>+15 compress the size of the changes.</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>+16 This paragraph contains text that is outdated. It will be deleted in the near future.</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>+17 It is important to spell check this document. On the other hand, a misspelled word isn't the end of the world. Nothing in the rest of this paragraph needs to be changed. Things can be added after it.</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>+25 This paragraph contains important new additions to this document.</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>+26 +</td>
</tr>
</tbody>
</table>

---

*Note: The changes are marked in red for deletions and green for additions.*
Diff

Longest Common Subsequence

https://en.wikipedia.org/wiki/Longest_common_subsequence_problem
Old school

- Why reinvent the wheel?
- Interest point detect/describe
- Match
- Works! 84-92% accuracy
- Good enough
- Awesome!
But there’s a catch

**Performance**

Movie is 200K frames
200K x 200K frame compares = 40B frame compares
64float descriptor at 8TFLOPs = 155 DAYS to compare with bruteforce :) 
24GB just to store the descriptor, not talking about any kd-trees etc
Comes to the rescue

Frequency Domain Descriptor

- Invented here at VideoGorillas
- DCT around interest point
- Frequencies as 160bit vector
- Hamming distance = SUPER fast
- we even bothered to patent it US20130243341A1

```c
int hamming_distance(uint64 x, uint64 long y) {
    return popcount(x ^ y);
}
```
theatrical 3000 frames

9M frame video LCS table

directors 3000 frames
Netflix - Edit Decision Reverse Engineering

- 300K frames edit from Orson
- 8.5M frames of 4K scans
- No Edit Decisions (EDL)
- We have a binary and a bunch of lines of code, decompile binary extract source code, reverse engineer Makefile
- 2.5T (trillion) frames to compare
- Run diff on 13K files
- Manually 5% in 9 months
## CPU vs GPU

### Performance

Rent 25 64 core servers for 24 hours
OR
Rewrite in OpenCL

<table>
<thead>
<tr>
<th>CPU</th>
<th>SSE asm</th>
<th>GTX1070 OpenCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPS</td>
<td>hours</td>
<td>days</td>
</tr>
<tr>
<td>18K</td>
<td>39351</td>
<td>1639</td>
</tr>
<tr>
<td>1.1M</td>
<td>614</td>
<td>25</td>
</tr>
<tr>
<td>12M</td>
<td>59</td>
<td>2.5</td>
</tr>
</tbody>
</table>
What’s next?

- Super resolution
- step1 infer 10% of a movie from 90%
- step2 infer 90% of a movie from 10%
- step3 super res 100%
Alex Zhukov

VideoGorillas