Empowering Visual Search
by CPUs, GPUs, and Compressed Kernels

Cees Snoek
University of Amsterdam
The Netherlands

Quiz: Search the ‘class rooms’

A B C D

E F G H

How difficult is the problem?

Human vision consumes 50% brain power...

Van Essen, Science 1992
The science of labeling

Worldwide endeavor in naming visual information

Visual labeling in a nutshell

Labeling is compute intensive

Distributed ASCI super computer: priceless

Visual labeling anno 2004

Visualization by Jasper Schulte
Empowering visual search by CPU

**Semstraa, MM 2007**

<table>
<thead>
<tr>
<th>Number</th>
<th>Tilbury</th>
<th>London University</th>
<th>University of Lund</th>
<th>Delft University of Technology</th>
<th>University of Oxford</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>15.37</td>
<td>15.14</td>
<td>15.23</td>
<td>15.03</td>
<td>15.10</td>
</tr>
<tr>
<td>4</td>
<td>7.30</td>
<td>7.35</td>
<td>7.64</td>
<td>7.30</td>
<td>7.48</td>
</tr>
<tr>
<td>8</td>
<td>3.69</td>
<td>3.74</td>
<td>3.51</td>
<td>3.57</td>
<td>3.95</td>
</tr>
<tr>
<td>16</td>
<td>3.14</td>
<td>3.09</td>
<td>2.96</td>
<td>2.98</td>
<td>3.09</td>
</tr>
<tr>
<td>24</td>
<td>1.98</td>
<td>2.05</td>
<td>2.08</td>
<td>2.08</td>
<td>2.13</td>
</tr>
<tr>
<td>40</td>
<td>1.87</td>
<td>1.98</td>
<td>2.06</td>
<td>2.06</td>
<td>2.13</td>
</tr>
<tr>
<td>56</td>
<td>1.96</td>
<td>1.94</td>
<td>2.08</td>
<td>2.08</td>
<td>2.13</td>
</tr>
<tr>
<td>60</td>
<td>1.87</td>
<td>1.98</td>
<td>2.13</td>
<td>2.13</td>
<td>2.13</td>
</tr>
</tbody>
</table>

Processing 184h of video: 250 compute days reduced to less than 60 hours

Visual labeling anno 2008

1. Trainen
   - Encode
   - Reduce
   - Learn

2. Testen
   - Encode
   - Reduce
   - Label

Van de Sande, PAMI 2010

We added color

David Lowe, ICCV 1999

Encoding simple cells with SIFT

Van de Sande, TMM 2011

Empowering visual search by GPU

![Graph showing performance comparison between CPU and GPU]
Exploit speed-up for more analysis

Analyzing 10x more video frames results in a relative accuracy gain of 29%

Visual labeling complexity

The number of images will grow forever

Kernel properties

Additivity
\[ K(p, q) = K(p_1, q_1) + \cdots + K(p_D, q_D) \]

Homogeneity
\[ K(c \cdot p, c \cdot q) = c^2 K(p, q) \]

Selecting and weighting dimensions

For additive kernels all dimensions are equal
\[ K_{X, Y} = 1 \cdot k_{X, Y} + \cdots + 1 \cdot k_{X, Y} \]

We introduce scaling factor \( c \):
\[ K_{X, Y} = c_1 \cdot k_{X, Y} + \cdots + c_D \cdot k_{X, Y} = c^T \cdot k_{X, Y} \]

Kernel reduction as convex optimization problem
\[
\arg\min_k \langle k_{X, Y} - c^T \cdot k_{X, Y} \rangle^2 + \lambda \| c \|_1
\]
Kernel error for various $\lambda$

*Similar accuracy with a 45-85% smaller size.*

Selected kernel dimensions

Note: descriptors originally dense sampled

International competition

NIST TRECVID Benchmark

Promote progress in video retrieval research

Open data, tasks, evaluation and innovation

Are we making progress?

MediaMill team

1000+ others
Performance doubled in just 3 years

SoJware licensed by UvA spin-off Euvision Technologies

Conclusion

For visual search it pays to be strong and smart.

DAS infrastructure empowers progress.

www.oesnnoek.info