Searching for the Perfect Picture

Dr Gerald Friedland
Jaeyoung Choi

Scientia

SEARCHING FOR THE PERFECT PICTURE

Currently, searching for images and videos on the Internet is less than ideal, with searches often providing completely irrelevant results. **Dr Gerald Friedland, Jaeyoung Choi** and their team at the International Computer Science Institute in Berkeley are working to develop better tools for image and video-based searches, to help researchers extract much more useful information.

Success in the current environment is increasingly dependent on Internet access – we find our jobs online, we pay our bills online, we identify that strange problem with our washing machine using information posted online. The Internet provides a vast sea of information about every topic imaginable – we can spend days learning about anything from auto repair to ancient Greece, from the history of zebra crossings to the health benefits of Zinc.

Yet all of this information is useless if we cannot find what we are looking for – the person looking for instructions to repair her car will be immensely irritated if all she can find are webpages about zebras. Thus, the ability to search and index online information is a necessity, and being able to do this efficiently and accurately has made search engines such as Yahoo and Google into multi-billion dollar corporations.

The complexity of the task means that search engines moved away from simple word matching long ago. Instead, the majority now balance information drawn from multiple sources – not only the text of the entire site being indexed but also the quality of other sites that link to it, and even the location and search history of the person typing the query. The outcome of this is that online searching is immensely more effective than it was a mere decade ago – to the point where we take relevant results for granted.

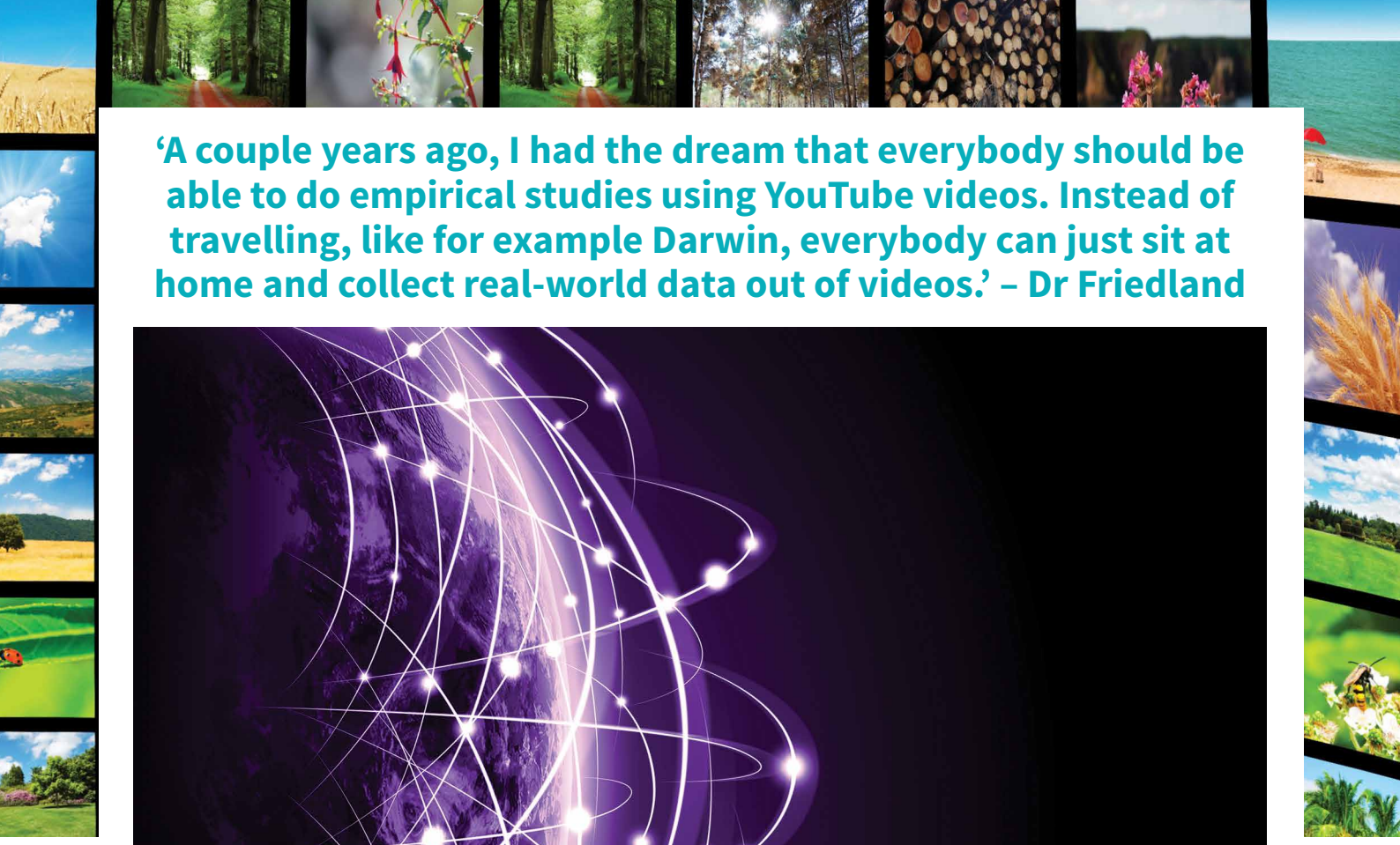
Part of this rapid increase in search relevance is due to the medium itself – the majority of the Internet is made up of text: words, sentences and paragraphs, just like this article. Text is very useful for computer analysis because it is relatively straightforward – it follows a well-organised set of rules (which we call 'grammar') and the meaning is relatively independent of the appearance (an 'a' is an 'a' regardless of whether it is written in Times New Roman or Comic Sans). This meant that search engine companies and their respective experts could focus on programming software to understand the meaning of text without worrying too much about extraneous problems.

Now, however, the Internet is rapidly filling up with visual media – there are hundreds of videos uploaded to YouTube every minute, and over one hundred pictures appear on Flickr in the same amount of time. A huge amount of information is now available in the form of film and images, yet it is exceptionally difficult to categorise.

This difficulty is predominantly due to the complexity of visual media – there is no 'grammar' for images, and no requirement that certain parts of the image be located and related in certain ways. Similarly, the same object can look very different in different images – a tree from the top and the side look completely unrelated, yet it is the same thing. This is not a problem for human viewers, as we excel in making conclusions based on limited data, but it is a nightmare for software systems.

The usual workaround for problems such as these falls under the lump term of 'big data' and 'machine learning'. Software can be written such that it will flexibly judge different pieces of a large data set, deciding for itself how to balance out the competing information. By 'training' the software on large amounts of known data, it will essentially write its own rules to analyse unknown data – rules that are often completely unlike those a human programmer would think of.

Machine learning approaches allow complex and ambiguous problems to be solved with ease, provided that there is sufficient data available at the beginning to train the system. They are, naturally, being employed in the field of image search.



‘A couple years ago, I had the dream that everybody should be able to do empirical studies using YouTube videos. Instead of travelling, like for example Darwin, everybody can just sit at home and collect real-world data out of videos.’ – Dr Friedland

Rounding Up Data

Thus, the development of an efficient image search program requires two main things – it needs clever engineers and scientists to program the software, and it needs a lot of images and videos to train the software. The SMASH program, led by Dr Gerald Friedland and Jaeyoung Choi at the International Computer Science Institute, sets out to provide both of these requirements by helping industry giants and computer science students to team up.

Both researchers have long experience in the field of computer science studies, and were particularly interested by the potential for using the large data sets provided by online image and video-sharing sites. ‘A couple of years ago, I had the dream that everybody should be able to do empirical studies using YouTube videos,’ says Dr Friedland. ‘Instead of travelling around the world, like, for example Darwin, everybody can just sit at home and collect real-world data out of videos.’

This was, however, very difficult. YouTube videos are searchable using categories and keywords, both text-based functions, and both reliant on the uploader including enough accurate information to make a correct assessment of the video content. This is possible in principle, but is not very reliable in practice. ‘The problem is, YouTube and the like don’t really allow searching for videos with the accuracy required for scientific research,’ explains Dr Friedland. ‘For example, while there are tens of millions of cat videos, just try to find cats with a certain attribute doing a certain thing. So “white cat deaf” only results in ONE relevant video to the topic and one lecture that explains the scientific correlation between white blue eyes and deafness in cats. In other words, we need much better tools for search.’

Developing these tools required developing the right set of data to work with, and this is where Dr Friedland and Choi began to work on what would eventually be known as the Multimedia Commons project.

Multimedia Commons

To improve research in the field of image and video, Dr Friedland, Choi and their colleagues formed a group called Multimedia Commons. Through this initiative, the team is heavily involved in improving a publicly available dataset known as the Yahoo-Flickr Creative Commons 100 Million dataset, usually referred to as YFCC100M for simplicity. This dataset contains the metadata of about 99.2 million photos and 800 thousand videos – details on where the image was taken, who took it, what camera was used, etc. Although this is useful, the dataset does not contain any of the actual images themselves, which excludes those researchers who are interested in extracting information from the image itself rather than the metadata.

To make the information more useful, the Multimedia Commons program has collaborated with Yahoo to provide all of the original images and videos. This improved dataset has then been enhanced by adding further information, such as visual features, image information, and other annotations. The enhanced dataset has already been used in several different ways by organisations affiliated with the SMASH/Multimedia Commons groups.

One of these uses is known as the MediaEval Placing Task – a competition organised by the team in which contestants must develop software that can estimate where a photo has been taken, based on the content of

the image and the additional data. Each contestant is provided with 5 million photos to help train their machine learning software, they are then graded on the identification of a set of 1.5 million images. This competition helps to develop the next wave of image recognition and search software, bringing together multiple different approaches with a common goal. Competitions of this sort also lead to various entertaining discoveries, such as Dr Friedland's recent finding that the ability to locate a photo can be reduced through the use of simple Instagram filters – a definite plus for those worried about the privacy of the photos that they upload to the Internet.

Data Analysis

Working through and analysing thousands of images requires slightly more power than the average home computer or laptop can provide. To speed up their work, many researchers in the field of machine learning rent computer time from a central server farm as they need it. This idea is actually quite old (having been behind the terminal/mainframe approaches of the first computers) but has been reborn under the term 'cloud computing' – researchers do not need to own a blindingly fast computer, they just need to rent it from someone who does.

One of the biggest companies in the cloud computing space is Amazon, who provide processing time to companies and institutions through their Amazon Web Services subsidiary. This subsidiary has grown from a small start-up into a billion-dollar business used around the world. Amazon Web Services is currently collaborating with the SMASH/Multimedia Commons group to provide free hosting of the vast amounts of data required by the scientists. This can be downloaded for free if required, but many researchers choose to use cloud-based streaming and storage processes to allow them to access only the required data as they go. This saves on download bandwidth and storage space (as 100 million pictures take up quite a few hard drives) and speeds up the process of analysing the data significantly.

Dr Friedland, Choi and the team have further simplified life for their collaborators by developing a series of software tools which would allow the dataset to be analysed. 'We built tools which allowed simple interfacing between Jupyter Notebook, a beginner's data science tool, and image search in the cloud,' says Dr Friedland. 'An example tool allows you to predict the location in which an image was taken based on the similarity to other images.' These tools are designed to be launched easily on Amazon Web Services, and so can be used in combination with Amazon's existing machine learning framework and infrastructure. This essentially allows even the worst-equipped computer scientist to develop their own multimedia search programs.

The Future of Search

The amount of information available to us is currently increasing at an astonishing rate – news articles, books, pictures and videos are appearing online every millisecond. Yet none of this is useful if we cannot identify the information we need amongst the vast sea of irrelevant results. In particular, the current status of image and video-based searches is less than ideal – results are often obscure or not related to the subject at hand. The work of groups such as Multimedia Commons are thus vital in providing the groundwork for these image searches, essentially setting up the laboratory in which these new discoveries will be made.

Perhaps, one day, you will be able to confidently ask your computer to find that picture of your Aunt Gladys by the pool in summer 2017 – and be confident that you will get the correct picture back. When that day arrives, it will be in no small part due to the actions of researchers such as Dr Gerald Friedland and Jaeyoung Choi.

Meet the researchers

Dr Gerald Friedland
Principal Data Scientist
Lawrence Livermore National Lab
Livermore, CA
USA

Jaeyoung Choi
Staff Researcher
Audio and Multimedia Group
International Computer Science Institute
Berkeley, California
USA

Dr Gerald Friedland's research career has stretched across the globe, having begun with a PhD from the Freie Universität Berlin in Germany and progressing over his role as Director of the Audio and Multimedia lab at the International Computer Science Institute to his current role as the Principal Data Scientist at Lawrence Livermore National labs. He is also an adjunct faculty member at the EECS department of the University of California, Berkeley. With over 200 published articles in conferences, journals, and books, and recipient of a number of awards, Dr Friedland is a leader in the field of multimedia search and processing.

CONTACT

E: fractor@icsi.berkeley.edu
W: http://www1.icsi.berkeley.edu/~fractor/homepage/Welcome.html

Jaeyoung Choi is currently a staff researcher at the Audio and Multimedia lab of the International Computer Science Institute in Berkeley, California. Much of his research to date has focused on the extraction of useful information from multimedia and in the management of online privacy due to image and information sharing. His years of working on this topic have led to a number of publications and awards. He is currently also conducting research towards attainment of a PhD degree from Delft University of Technology in the Netherlands.

CONTACT

E: jaeyoung@icsi.berkeley.edu
W: http://jaeyoungchoi.com/

KEY COLLABORATORS

Julia Bernd, International Computer Science Institute, USA
Damian Borth, German Research Center for Artificial Intelligence (DFKI), Germany
Carmen Carrano, Lawrence Livermore National Laboratory, USA
Benjamin Elizalde, Carnegie Mellon University, USA
Ariel Gold, US Department of Transportation
Luke Gottlieb, Synopsis, Inc., USA, & ICSI
Adam Janin, Mod9 Technologies
Martha Larson, TU Delft & MediaEval Benchmarking Initiative
Karl Ni, In-Q-Tel, USA, & LLNL
Doug Poland, Lawrence Livermore National Laboratory
David A. Shamma, FXPal, USA
KD Singh, Amazon Web Services
Bart Thomee, Alphabet, Inc
Jed Sundwall, Amazon Web Services
Joseph Spisak, Amazon Web Services

FUNDING

This work was graciously supported by AWS Programs for Research and Education. We also thank the AWS Public Dataset program for hosting the Multimedia Commons dataset for public use. Our work is also partially supported by a collaborative LDRD led by Lawrence Livermore National Laboratory (U.S. Dept. of Energy contract DE-AC52-07NA27344) and by National Science Foundation Grant No. CNS 1514509.

ICSI INTERNATIONAL COMPUTER SCIENCE INSTITUTE
Berkeley UNIVERSITY OF CALIFORNIA
Lawrence Livermore National Laboratory