Real-time Face Recognition in Live Video Streams

University of Leicester, UK
(in partnership with ARM, VMS Ltd, and LITSO)

Introduction

Face recognition is something everyone has thought about at some point in their lives. Face recognition is what we as humans can perform with ease. Face recognition allows us to interact with people and essentially recognise them. But unlike computers our memory of faces is limited. There are applications, however, that do require fast and efficient facial recognition capabilities. An example of these are large-scale security systems in public places. In 2016 there were about 206,800 passengers arriving every day in Heathrow. Screening this volume on the basis of human experts alone is expensive, and thus employing automated face recognition and identity matching technologies is a possible remedy to tackle the problem.

Another sphere of immediate demands for such a technology is the market of surveillance systems in the UK. The UK market for security systems was ca. £2bn in 2010 alone (BSRIA report). The UK is recognised as the most surveyed country in Europe with an estimated 5.9m CCTV cameras deployed. With a modest 8 Mbps one camera generates 3,6GB of HD data per hour, leading to Petabytes of visual data nationwide daily. Online Spy Shop reports that in 2016 the average Briton is caught 70 times a day. Yet, in 2008 the Met Police reported that only one crime was solved for every 1,000 cameras. The 2015/2016 Annual Surveillance Camera Commissioner Report suggests that hard empirical evidence to support the existence of surveillance in our towns and cities is lacking. It states, however, that if processing resources were there then CCTV could be extremely valuable (cf p. 14). At the same time, Local Authorities spent ca. £100M on salaries of CCTV in 4 years, and these resources are stretched. To alleviate costs and increase efficiency automated AI systems for security and surveillance are urgently required.

Specific problem

Computer vision systems are at a key point in their evolution. Their importance stretches across robotics and autonomous vehicles to automated approaches in health care. Research and development in this field has progressed over the last 60 years. Key themes in the literature include perceptrons, statistical theory of pattern classification, support vector machines, and neural networks.

Recently, extremely capable new technologies arrived: Convolutional Neural Networks, Deep Learning, Deep Belief, Random Field, and residual Neural Networks. They show remarkable performance in benchmark tests and applications. Despite this success several obstacles impede further progress and limit the impact of these discoveries on society. These are:

1. Computational costs.
2. Autonomy and energy consumption
3. Training of these advanced systems requires large data set of samples that is not necessarily available for all objects of interest and applications.

Solving all or a part of these limitations will enable immediate and broad access of the technology to society.

ARM, a UK-based company, has developed, in partnership with the University of Leicester, a prototype of a low-cost and a low-energy system, featuring image representation with Histograms of Oriented Gradients at multiple spatial scales and linear Support Vector Machine-type classifiers for object detection. The system has been fully implemented in the hardware; this allows processing of high-resolution images and video streams in near real-time speed of approximately 30 frames per second. The system, however, has been trained to ***detect*** faces. The task now is to ***develop face recognition capabilities***.

Approach and work plan

On the first day of the school participating students will be provided with a tutorial on the particular system at hands and on the basics of Machine Learning Theory that is relevant for the project. Remaining days will be devoted to independent yet supervised take on the problem. In the process of solving this task the students will obtain first-hands experience in training a real functional object detection system. They will also acquire practical data mining and data

analysis skills and will experiment with measure concentration effects in high dimensions. To minimize development time, we will begin with legacy face recognition systems such as e.g. VGG face [4] for real-time automated tagging of facial images in real time. At the end of the project it is expected the students will develop and present a working system capable of matching facial data from real-time video streams against a given list of identities. Hardware, developed by ARM in partnership with the University of Leicester, for detecting and streaming facial crops will be provided by project facilitators.

Recommended Literature

[1] R. O. Duda, G. Stork, P. E. Hart. Pattern classification and scene analysis, Wiley, 2000.
[2] V. Vapnik. The Nature of Statistical Learning Theory, Springer, 2000.
[3] N. Dalal, B. Triggs. Histograms of Oriented Gradients for Human Detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 1063-6919.
[4] http://www.robots.ox.ac.uk/~vgg/software/vgg_face/