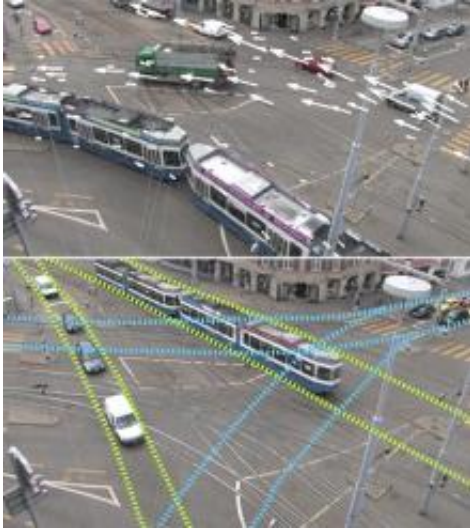# Smart computer learns from video

June 23 2010

The computer recognises the spatial and temporal patterns of sequences of activities in road traffic. (Image: D. Küttel / ETH Zurich)

Swiss researchers have written a computer programme that is able to analyse temporal and spatial patterns of moving objects, and on top of that is capable of learning. This would be a significant aid in traffic monitoring.

The number ten tram crosses the carriageway, makes a sharp bend to the right, and stops in front of the Maschinenlaboratorium building. At the same time, cars roaring down Universitätsstrasse are forced to stop, students scurry across the zebra crossing and the number six tram heading towards the zoo comes around the corner. A typical traffic scene

in the city of Zurich. The scene repeats itself in a regular cycle, and spatial and temporal patterns can be recognised.

For a person armed with a stopwatch and something to write with, it wouldn't be too difficult to track and analyse these patterns. For a computer, however, it is a formidable challenge to collect all the data and analyse such a scenario and on top of that to memorise the patterns too.

Smart algorithm learns patterns

But ETH researchers Daniel Küttel and Michael Breitenstein, together with professors Luc Van Gool and Vittorio Ferrari from the Institute of Image Processing have now developed a computer code - an algorithm - that does precisely this. The software is able to analyse street scenes like this from video images, and map the spatial and temporal patterns that characterise the various road users. The computer can recognise which tram is going past and when, how many cars pass through in its 'tracks', and other such details. The computer also registers any deviations from a normal situation.

To create this programme, the researchers mounted cameras at a number of junctions around the city of Zurich and recorded hours of video clips. The computer then analysed these video sequences and automatically, i.e. without intervention by the programmers, established rules governing the flow of traffic. The computer had to spend about a day working out the calculations for each hour of video footage that had been recorded. Once the machine had 'learned' the standard patterns, however, it was then able to interpret the video recordings in real time.

Computer generates tram timetable

The researchers used comparisons with the tram timetable to ascertain that their programme works extremely accurately. To test the programme's function, they viewed the timetable on the internet and simultaneously monitored how the computer analysed the traffic scenes. These automatic analyses corresponded exactly with the timetable, to the minute, confirming beyond any doubt that the machine was analysing the video data correctly.

Theory was the hard part

What sounds simple was actually tough programming work. It took lead author Daniel Küttel, a doctoral candidate of Vittorio Ferrari, nine months to programme the algorithm. "The hardest part of it was processing the theory behind it", he says. For the researchers, the first priority was to test the concept. Evidently, they did this so well that at the IEEE Conference on Computer Vision and Pattern Recognition held this June in San Francisco, the IEEE professional organisation included this work in the top five per cent of 1,200 contributions submitted.

Although an application has yet to be determined, the project is already complete. Ferrari believes, however, that the automatic interpretation of dynamic camera images could help with the monitoring of traffic from a traffic management control centre. The advantage: one person could be in charge of a number of monitors simultaneously, since the computer would immediately detect anomalies, for example, if the flow of traffic comes to a halt or a [car](link) goes the wrong way down a one-way street. The person manning the monitor could then decide whether any action needed to be taken in the specific situation.

Automatic image recognition

In their next project, the researchers will attempt to train the computer to

recognise visual concepts, to trawl internet pages in search of specific images and automatically find the correct image, including those that are not labelled accordingly. If the computer can be taught, for example, what Roger Federer looks like, it should then automatically find images of Federer with minimal human intervention, including when the text caption says something completely different, or nothing at all.

  More information: Kuettel D, Breitenstein MD, Van Gool L, Ferrari V. What's going on? Discovering Spatio-Temporal Dependencies in Dynamic Scenes. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10), San Francisco, 2010.
www.vision.ee.ethz.ch/~calvin/ … s/kuettel-cvpr10.pdf

Provided by ETH Zurich