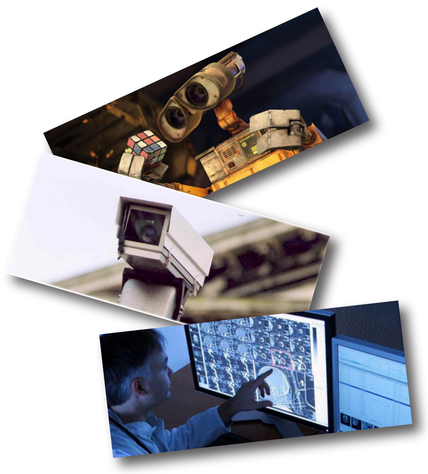
UK ICT Pioneers Competition

Computer Vision-based Event Detection

Jing Wang
University of Huddersfield

Introduction



Computer vision is concerned with the theory and practices for building artificial systems to obtain information from images and video footages.

Video recordings provide rich data on dynamic events occurred over a period of time.

Video event detection is a hotly-pursued computer vision research area for automatically detecting and interpreting real-world activities.

Challenges

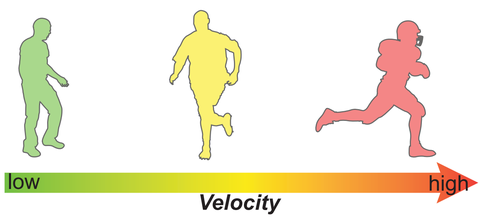
The large variations on the definition of video events and the low efficiency for most of today's event detection techniques are still posing great challenges to their wider adoption.

The semantic of an "event" in a video is ambiguous;

The boundary between an "event" signal and its "background" noise is often inexplicit;

The uncertainty of durations for video events;

The conceptual and contextual "understanding" of a video event is still a hugely challenging task requiring major advancements from other research domains such as machine and artificial intelligence.



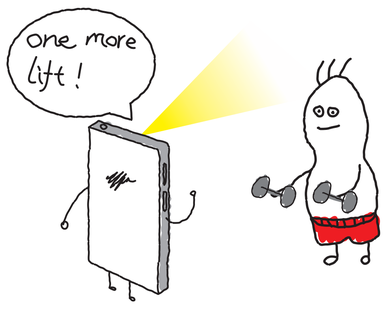
Applications

Video Retrieval

Digital library management
Sports video analysis
Intelligent auto-editing

Surveillance Systems

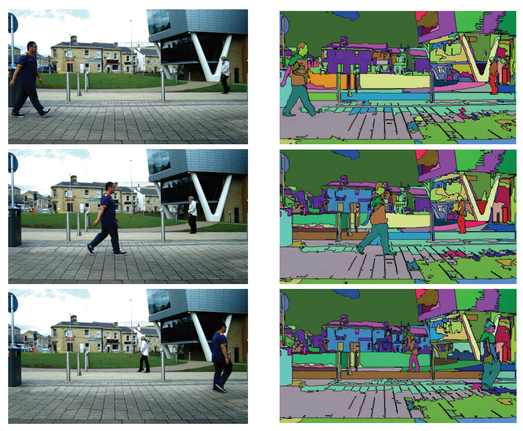
Graffiti detection
Vandalism radar



Pervasive HCI

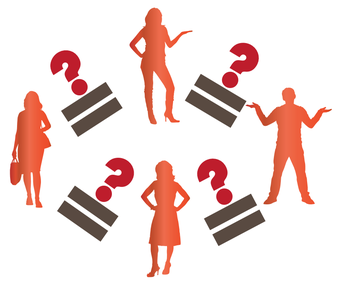
Personalised health care
E-and-mobile learning

Main Contributions



An innovative video data representation and abstraction technique:

Representing spatial and temporal video information as 3D volumetric models

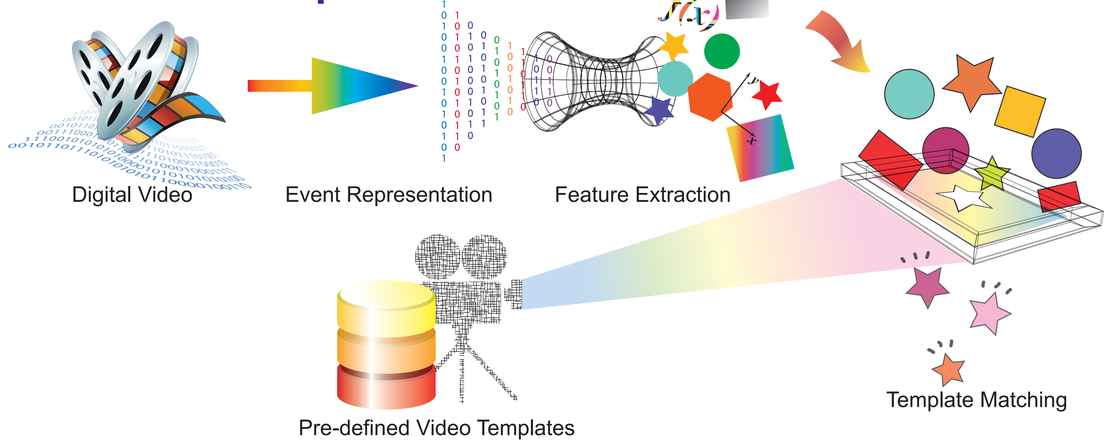


$$d_{N,C,P}(T,V;l)=d_{N,C}(T,V;l)+p*d_P(T,V;l)$$

An efficient video event template matching algorithm:

Detecting video events by recognising shapes and distributions of those 3D models effectively

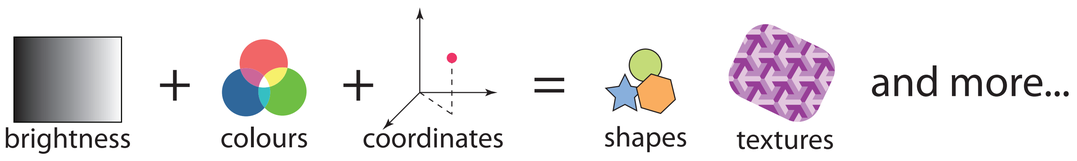
Process Pipeline



Video Event Representation

Videos are represented by the Spatio-temporal Volume (STV) data structure, which involves the spatial and temporal information extracted from video footages.

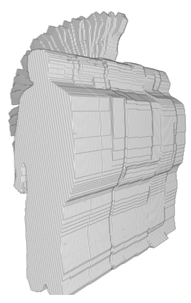
Video Feature Extraction



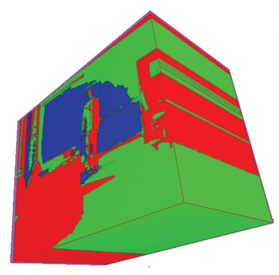
Video event information can be encapsulated into corresponding 3D shapes abstracted from STV data by using an innovative image segmentation technique developed in this research.

Template Matching

Based on the segmented outputs, video event information are encapsulated into corresponding 3D shapes, which transforms the event detection tasks into 3D shape matching operations. Compared with other existing matching processes, the algorithm demonstrates stronger robustness against complex real-world settings.

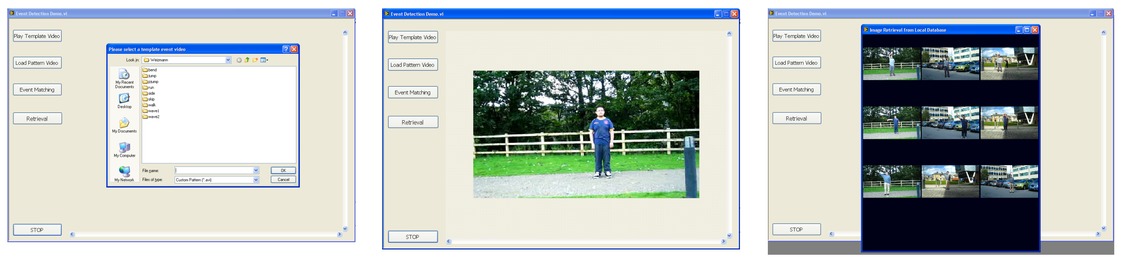


Template



STV Shape Features

System Prototype



Test Results

The system has been tested on public databases and footages recorded in the University of Huddersfield campus containing various events. Compared with existing approaches, this algorithm shows improved Recall-Precision rate and faster operational speed.

Conclusion and Future Works

This research has introduced the so-called Spatio-temporal Volume technique into the 3D shape matching domain for fascinating video event detection that can be applied in many digital video management and analysis applications.

In the future, the algorithm can be extended to other platforms such as mobile and distributed devices. The system performances can also be further improved through employing hardware acceleration and data parallelism enabled by evolving computer technologies.

Funded by
University of HUDDERSFIELD