Real-time Video Analysis

Each day in 2019, around 2,000 petabytes of video data are generated by security cameras around the world, up from 500 petabytes per day in 2015. This is equivalent to 75 million users streaming an hour's HDTV simultaneously[1]. This dramatic growth is driving the demand for more efficient methods of analysing video data.

New video analytics from the University of Warwick can outperform other methods, whilst requiring less processing and training time, and less storage. We are now seeking to licence our algorithm to solution providers, system integrators and companies developing video analytics.

Dr Victor Sanchez and his team at the University of Warwick have designed a new method of feature descriptor for video analytics that encodes the motion information of a **Spatio-Temporal** support region into a low-dimensional **Binary** string (STB).

Applications

Our algorithm is able to address these needs in video captured in real environments.

> classification
> action recognition
> object recognition
> video surveillance
> monitoring
> abnormal event detection

Advantages

> low computational times and reduced memory and storage requirements

> increased descriptive power through encoded information from two motion sources

> much shorter training times compared to current systems

> suitable for real-time applications and in devices with low-computational capacity

> significantly outperforms other binary descriptors including 3D-FREAK, 3D-BRISK, 3D-ORB & 3D-BRIEF

Ref: 1. IHS Markit, January 26 2016

For further information contact: Dr Shum Prakash, Warwick Ventures
Te: +44 (0) 24 7657 4145 Email: s.prakash@warwick.ac.uk

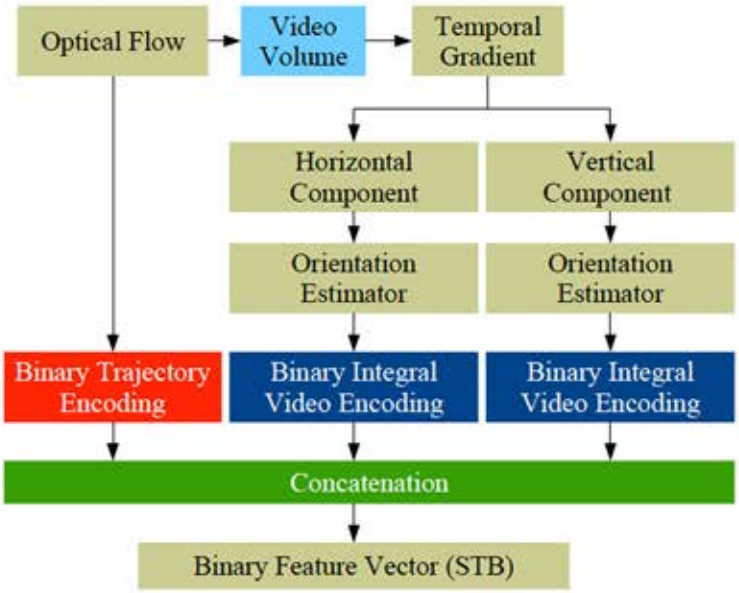
Solution Overview

The encoded motion information is obtained from two motion sources: optical flow and temporal gradients, which provide rich motion information by considering pixel intensity changes to create a new data space that disregards the background.

The STB Descriptor

The first binary string represents the video volume's motion information extracted from the optical flow – Binary Trajectory Encoding.

The second and third binary strings represent the independent horizontal and vertical motion components of the video volume's temporal gradients – Binary Integral Video Encoding (BIVE). The motion components are encoded using wavelet-based patterns with large regions.



Feature Descriptor Performance Evaluation and Comparisons

Based on extensive evaluations for action recognition in videos, BIVE attains the highest correct classification rate (CCR) among several binary feature descriptors.

BIVE significantly outperforms 3D-FREAK, 3D-BRISK, 3D-ORB and 3D-BRIEF binary descriptors. This confirms the advantages of using wavelet-based patterns and relatively large regions to encode the temporal gradients of video volumes. Use of the Integral Video technique was observed to speed up the encoding process by approximately 40× compared to the case of not using it.

The STB descriptor, when using the Integral Video technique, is around 1200× faster than 3D-SIFT, and 200× faster than HOG. In terms of memory demands, feature vectors generated by STB are c.a. 830× more compact than those generated by 3DSIFT, and 30× more compact than those generated by HOG. When used for video analytics, for example action recognition, a machine learning system using the STB descriptor can be trained in hours, which is much faster than the training time required by common CNN-based systems, which may be in the order of several days.

Patent & Publication

International patent application number: PCT GB2018/058103 26 October 2018

R. Leyva, V. Sanchez, and C.-T. Li, "Fast Detection of Abnormal Events in Videos with Binary Features," Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1318-1322, April 2018, Calgary, Canada

For further information contact: Dr Shum Prakash, Warwick Ventures
Te: +44 (0) 24 7657 4145 Email: s.prakash@warwick.ac.uk