Abstract

The task of video anomaly detection is most challenging in defining the objectives as it is diversified in approaches and interpretation of the problem. The various anomaly detection algorithms address the challenges encountered due to the design of the algorithms and classification models used for anomaly detection. The computational time and complexity, accurate detection and localization play a major role in implementing the algorithms. The challenges such as data collection, data preparation, and evaluation strategies become equally important key issues to be considered by the computer vision community.

In crowded scenes, global abnormal events form unique and distinct motion characteristics, and category of anomalies. Such events require image-level detection rather than pixel-level localization with less complexity compared to local abnormal events. The features are generally computed at pixel-level or block-level and are of motion and appearance types. The pixel based approaches extract features from foreground pixels leads to heavy computational complexity. The block based approaches address the problem of heavy computations and achieve high detection rate for local anomalies but fail to achieve better localization. Such methods are suitable for detection of global abnormal events. Hierarchical block based approaches compensate between the computational complexity and localization problems.

The above discussed pixel based, block based and hierarchical block based methods use background modeling algorithms, detect either global abnormal events or local abnormal events, and use the traditional histogram of optical flow (HOF) motion descriptor. The use of background modeling algorithms extract foreground objects by reducing the computational burden, but their efficiency highly influences the detection accuracy. The traditional HOF descriptor fails to capture the local motion variation and structure information of local anomalies. The focus of this thesis work is to address these problems.

In the initial section, we present a survey of various video anomaly detection datasets. The description and categorization of these datasets are briefly explained.

In the next section, we address the drawbacks of existing pixel based and block based anomaly detection methods. We propose block based approaches viz Hierarchical block based feature extraction method, and Block based feature extraction method using context location and motion-rich spatio-temporal volumes (MRSTVs) to solve the problem of global abnormal events detection in crowded scenes. These methods use low-level optical flow motion features extracted from spatio-temporal volumes (STVs) to represent motion information of global abnormal events. We use one-class SVM to learn normal events and detect abnormal frames. A combined spatial and temporal post-processing

method is proposed to reduce false alarm rate and improve the detection results. The aim of these methods is to omit the use of background modeling algorithm and detect location-related anomalies. Motion-rich spatio-temporal volumes (MRSTVs) imply STVs belong to foreground objects. Context location is obtained as a byproduct of this method that helps to define unexpected/restricted motion area to detect location-related anomalies.

The proposed global abnormal events detection methods are combined to detect global and local abnormal events detection via hierarchical block based feature extraction. Extraction of block based features at three levels of hierarchy reduces the computational burden. This is achieved by using global analysis model for frame-level detection and local analysis model through hierarchical structure to achieve localization for local abnormal events. Further, global and local abnormal events are categorized. In all the above proposed methods, the focus was to address the computational burden and achieving reasonable localization accuracy for pixel based and block based approaches, adopting different models for hierarchical approaches, and detecting location-related anomalies.

In the final section, we propose a new 3D Spatially-Localized Histogram of Optical Flow (SL-HOF) motion descriptor for detecting local abnormal events. The 3D SL-HOF descriptor has more discriminative power than the traditional HOF motion descriptor which is widely used to represent motion characteristics of abnormal events. The 3D SL-HOF descriptor utilizes motion magnitude and orientation information from optical flow map more efficiently to capture local motion variation and structure information of foreground objects. Appearance features computed using 3D HOG descriptor and combined with motion features. Motion-rich STVs extract foreground objects whereas context location helps to detect abnormal crowd movement in an unexpected region. Finally, One-Class SVM classifier is adopted to train the normal behaviour and detect abnormality during testing.

The proposed 3D SL-HOF descriptor is used for spatio-temporal trajectory analysis by extracting motion and appearance features around the detected keypoints. The computationally efficient oriented FAST and rotated BRIEF (ORB) keypoint detector is used to detect keypoints and tracked using KLT algorithm. A new 3D SL-HOF motion descriptor and 3D HOG appearance descriptor are used to extract features around the detected and tracked keypoints. Bags-of-words model is used to encode these features. One-Class SVM classifier is used to train the normal behaviour and detect abnormal behaviour. The use of 3D SL-HOF motion descriptor is proven to be efficient in representing motion features with improved detection accuracy compared to the HOF descriptor.