Chapter 95
An Empirical Study for Human Behavior Analysis

Jia Lu
Auckland University of Technology, New Zealand

Jun Shen
Auckland University of Technology, New Zealand

Wei Qi Yan
Auckland University of Technology, New Zealand

Boris Bačić
Auckland University of Technology, New Zealand

ABSTRACT

This paper presents an empirical study for human behavior analysis based on three distinct feature extraction techniques: Histograms of Oriented Gradients (HOG), Local Binary Pattern (LBP) and Scale Invariant Local Ternary Pattern (SILTP). The utilised public videos representing spatio-temporal problem area of investigation include INRIA person detection and Weizmann pedestrian activity datasets. For INRIA dataset, both LBP and HOG were able to eliminate redundant video data and show human-intelligible feature visualisation of extracted features required for classification tasks. However, for Weizmann dataset only HOG feature extraction was found to work well with classifying five selected activities/exercises (walking, running, skipping, jumping and jacking).

INTRODUCTION

Rapid increase of surveillance cameras capturing human activities, processing of events of interest in a scene is still considered as an on-going challenge. In general, for loosely defend 'intelligent' surveillance, a video acquired from networked cameras should be processed in such way that it could provide meaningful information that we could use and further process if needed.

As an active research area in intelligent surveillance, event recognition mining and reasoning can improve the accessibility and reusability for a large number of media collections (Maryam & Reza, 2012).

Event recognition, which has a wide range of applications, can also be utilized to identify particular events as a function to find out abnormal human behaviours (Chen & Zhang, 2006). Event in our real world can be defined as occurrence happened in a determinable space and time (Popoola & Wang, 2012).

To reduce human labour involved in visual detection and events/activities recognition from video streams and to label a surveillance system as 'intelligent' at present time, it is necessary to achieve a degree of automation in information processing from videos. To achieve these objectives, we aim to design computer infrastructures for monitoring our environments in 24 hours per day, seven days one week.

Despite growing number of surveillance videos from various sources (analogue and digital) that have been analysed for decades, object, event, activity detection and recognition have been well investigated, this problem still remains. In recent years, there is also growing attention in our research community for data processing from the surveillance with focus on how to detect and recognize objects with spatio-temporal relationships (Krumm, et al., 2000) (Siebel & Maybank, 2002). The importance for this research is at semantic understanding of human behaviour. Automated human behaviour recognition plays a pivotal role in storing video streams into a database for further analysis. In literature, there are two categories of automated event analysis, namely spatio-temporal models and parametric models (Rui & Anandan, 2000) (Cutler & Davis, 2000).

In this paper, we assert that in surveillance systems there is no single best approach. In addition, the process of Feature Extraction (FE) as video data reduction and intermediate visualization of data processing stages may provide additional insight into specific surveillance problem or broader context that may inform further advancements in this field.

This paper is organized as follows: Section 2 includes background and related work of person and event detection in surveillance, ANN and feature selection techniques; Section 3 will describe utilized datasets, application of existing feature extraction techniques with comments on perceived visualisation of intermediate data processing for classification tasks; Section 4 reports on the results achieved; and the conclusions and future work will be summarised in Section 5.

BACKGROUND AND RELATED WORK

In surveillance, we describe an event in six facets, namely, What, When, Who, Why, Where and How (5W1H) that could be generalized to feature any surveillance events (Westermann & Jain, 2007). Aligned with studies in video mining and video retrieval (Dai, Zhang, & Li, 2006) (Geetha & Narayanan, 2010), events are regarded to consist of these six 5W1H major components in event recognition and modelling (Xie, Sundaram, & Campbell, 2008). In computing, visual event represents an action or occurrence that could be quantified and recognised by (computing) machine. Similarly, the definition of an event in this paper is the occurrence of something at a particular time and at specific location. In order to facilitate a computer to record, index and arrange video events for users' post-analysis, the events have a number of attributes including ID, time, location and description. According to the attributes, an event is detected and classified into different classes from the videos in surveillance. Based on categories of an event, we can group the detected events as normal and abnormal ones. For example, Figure 1 shows a normal and abnormal event. Normally, a pedestrian should walk in standing position as in Figure 1(a) or when the walker/bystander falls down as in Figure 1(b) the abnormal event should be detected, and a surveillance alarm should be generated correspondingly and automatically.

Related Content

Effective Technique to Reduce the Dimension of Text Data
D.S. Guru, K. Swarnalatha, N. Vinay Kumarand Basavaraj S. Anami (2020). *International Journal of Computer Vision and Image Processing (pp. 67-85).*
www.irma-international.org/article/effective-technique-to-reduce-the-dimension-of-text-data/245670

3D Face Recognition Using Spatial Relations
Stefano Berretti, Alberto del Bimboand Pietro Pala (2018). *Computer Vision: Concepts, Methodologies, Tools, and Applications (pp. 679-706).*
www.irma-international.org/chapter/3d-face-recognition-using-spatial-relations/196974

A Blind 3D Watermarking Approach for 3D Mesh Using Clustering Based Methods
Mona M. Soliman, Aboul Ella Hassanienand Hoda M. Onsi (2013). *International Journal of Computer Vision and Image Processing (pp. 43-53).*
www.irma-international.org/article/a-blind-3d-watermarking-approach-for-3d-mesh-using-clustering-based-methods/87250

PAP/CAR-Based Water Erosion Analysis in the Oued Lakhdar Watershed: A GIS and Remote Sensing Approach
Rachid Laaouidi, Siham Roudani, Mouhcine Batchiand Zakariae Amhani (2025). *Remote Sensing and GIS Techniques in Hydrology (pp. 259-296).*
www.irma-international.org/chapter/papcar-based-water-erosion-analysis-in-the-oued-lakhdar-watershed/360450

The Role of Significant Digits of Fractal Dimension for the Measurement of Texture in SAR Images
Triloki Pant (2021). *International Journal of Computer Vision and Image Processing (pp. 80-96).*
www.irma-international.org/article/the-role-of-significant-digits-of-fractal-dimension-for-the-measurement-of-texture-in-sar-images/270877