Human Action Recognition Using Median Background and Max Pool Convolution with Nearest Neighbor

Bagavathi Lakshmi, Research and Development Centre, Bharathiar University, Coimbatore, India
S.Parthasarathy, Department of Computer Applications, Thiagarajar College of Engineering, Madurai, India

ABSTRACT

Discovering human activities on mobile devices is a challenging task for human action recognition. The ability of a device to recognize its user’s activity is important because it enables context-aware applications and behavior. Recently, machine learning algorithms have been increasingly used for human action recognition. During the past few years, principal component analysis and support vector machines is widely used for robust human activity recognition. However, with global dynamic tendency and complex tasks involved, this robust human activity recognition (HAR) results in error and complexity. To deal with this problem, a machine learning algorithm is proposed and explores its application on HAR. In this article, a Max Pool Convolution Neural Network based on Nearest Neighbor (MPCNN-NN) is proposed to perform efficient and effective HAR using smartphone sensors by exploiting the inherent characteristics. The MPCNN-NN framework for HAR consists of three steps. In the first step, for each activity, the features of interest or foreground frame are detected using Median Background Subtraction. The second step consists of organizing the features (i.e. postures) that represent the strongest generic discriminating features (i.e. postures) based on Max Pool. The third and the final step is the HAR based on Nearest Neighbor that postures which maximizes the probability. Experiments have been conducted to demonstrate the superiority of the proposed MPCNN-NN framework on human action dataset, KARD (Kinect Activity Recognition Dataset).

KEYWORDS

Convolution Neural Network, Human Action, Human Activity Recognition, Machine Learning Algorithm, Max Pool, Sensors

INTRODUCTION

One of the most widespread researches in computer vision field is Human Activity Recognition (HAR), because of its extensive significances. To name a few are, surveillance of video, entertainment through human and computer, health care analysis, social welfare and so on. The main purpose of HAR remains in examining and identifying human activities by machines. In spite of consequential efforts have been made in the past several years, identifying human action precisely is still a demanding problem.

Human activity recognition using smartphones benefits several applications. To name a few are, healthcare system, body fitness, smart home and so on. Instead of using smart phones that are invasive and necessitate additional cost, using smart phones integrated with different types of sensors are found to be of extensive use. In a previous research work by Chen et al. (2017), Coordinate Transformation
and Principal Component Analysis (CT-PCA) was investigated for robust human activity recognition with respect to direction, position.

The CT-PCA scheme addressed one of the most significant issues with respect to the effect of orientation variations in HAR based on smart phone. In addition to improve the performance of recognition, an efficient Online-independent Support Vector Machine (OSVM) algorithm was designed. Despite improvement found in recognition performance of the system, however, with complex tasks, representative features are not selected in a proper manner. This in turn results in error and therefore recognition accuracy for human activity recognition was said to be compromised. Potential solution is to introduce machine learning algorithms that extract representative features even for complex tasks, therefore addressing the error rate.

In Posture Tendency Descriptor (PTD) Yao et al., (2018) an intuitive dividing algorithm was first designed that efficiently divided the action sequence into various action snippets. The purpose of applying the intuitive dividing algorithm was the geometric structure within the activity was preserved, resulting in the improvement in the recognition accuracy. With the obtained action snippets, an interpretable and discriminative posture tendency descriptor (PTD) was built with the purpose of representing one action snippet.

Finally, multiple PTDs were integrated in a vertical and time-related order resulting in the human activity representation, therefore reducing the computational complexity involved in human activity recognition. Despite improvement addressed in terms of recognition accuracy, with the focus on the global dynamic tendency by measuring the covariance of the entire body joint locations, could increase the computational complexity or the computational time involved.

Potential solution is to classify only the key frames by applying the Nearest Neighbor with the confidence of the nearest frames, therefore minimize the computational time involved during human activity recognition. A comparative study on human activity recognition was carried out by Wang et al. (2016), with novel feature selection approach resulting in improvement in terms of time.

Most prevailing techniques make the unfeasible speculation that persons walk along a stable orientation or a presumed path. Previously, a sparse reconstruction based metric learning method was investigated by Lu et al. (2014), thereby reducing the reconstruction error. Scale Invariant Feature Transform (SIFT) was used by Moussa et al., (2015) for detecting interest points for HAR. Yet another trajectory-based representation was designed by Haiaim et al., (2015), to extract effective trajectories in realistic conditions. This in turn had positive impact on the recognition accuracy. However, anticipation with respect to trajectory remained unaddressed. To solve this issue, Anticipatory Temporal Conditional Random Field (ATCRF) was applied in a previous research work by Hema et al., (2016), resulting in the improvement in anticipation time.

An analysis of the previous studies showed that they required strong representative features to be operating at all times with global dynamic tendency; as a result, without strong representative features being extracted and with global dynamic tendency, computation time and errors increased. In this paper, we argue that it is possible to extract strong representative features using machine learning algorithm. In order to further address the global dynamic tendency, a framework is proposed that fetches the key frames by applying Nearest Neighbour for HAR. Experiments are performed and conducted a thorough analysis of the result to show that the proposed MPCNN-NN framework performs well in terms of both computation time and HAR accuracy.

The remainder of the paper is organized as follows. Section 2 presents an analysis of previous HAR methods for efficient activity recognition. Section 3 describes our initial motivations and design of Max Pool Convolution Neural Network based on Nearest Neighbor (MPCNN-NN) framework for HAR. Section 4 explains the experimental setup for evaluating the MPCNN-NN framework. Section 5 reports on the evaluation results for our framework and comparison made with similar other existing methods. Finally, Section 6 concludes with a summary.
Self-organizing Mobile Sensor Network: Distributed Topology Control Framework
Geunho Lee and Nak Young Chong (2011). *Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives* (pp. 558-583).
www.igi-global.com/chapter/self-organizing-mobile-sensor-network/54674?camid=4v1a

Generating Knowledge-Based System Generators: A Software Engineering Approach
www.igi-global.com/article/generating-knowledge-based-system-generators/38988?camid=4v1a