Neurocomputing 173 (2016) 54–63

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Recognizing human actions using novel space-time volume binary patterns ☆

Florian Baumann [a,*], Arne Ehlers [a], Bodo Rosenhahn [a], Jie Liao [b]

[a] Institut für Informationsverarbeitung, Leibniz Universität Hannover, Appelstraße 9a, 30167 Hannover, Germany
[b] University of Science and Technology of China, Jinzhai Rd, Hefei, Anhui 230027, China

ARTICLE INFO

Article history:
Received 13 June 2014
Received in revised form
4 February 2015
Accepted 9 March 2015
Available online 7 August 2015

Keywords:
Random forest
Action recognition
patterns
Volume local binary patterns
KTH
Weizman
IXMAS

ABSTRACT

In this paper, we propose a novel feature type, namely Motion Binary Pattern (MBP) and different computation strategies for the well-known Volume Local Binary Pattern (VLBP). MBPs are a combination of VLBPs and Optical Flow. By combining the benefit of both methods, a simple and efficient descriptor is constructed. Motion Binary Patterns are computed in the spatio-temporal domain while the motion in consecutive frames is described. Finally, a feature descriptor is constructed by a histogram computation. Volume Local Binary Patterns are a feature type to describe object characteristics in the spatio-temporal domain. But apart from the computation of such a pattern further steps are required to create a discriminative feature. These steps are evaluated in detail and the best strategy is shown. For MBPs and VLBPs, a Random Forest classifier is learned and applied to the task of human action recognition. The proposed novel feature type and VLBPs are evaluated on the well-known, publicly available KTH dataset, Weizman dataset and on the IXMAS dataset for multi-view action recognition. The results demonstrate challenging accuracies in comparison to state-of-the-art methods.

1. Introduction

In this work, we address the problem of recognizing human actions performed by a single person, e.g. boxing, clapping, waving, walking, running, and jogging. Human action recognition is a complex area of computer vision since static object characteristics, motion and time information have to be taken into account. Furthermore, actions are divided into human actions, human–human interactions, human–object interactions and group activities [1]. Due to environment variations such as moving backgrounds, different view points or occlusions the detection and classification of actions is even more difficult. Additionally, each actor has its own style of performing an action, leading to many variations in the subject's movement and a large intra-class variation [1,27].

This paper is mainly divided into two parts. In the first part, we propose an efficient novel feature type, namely Motion Binary Pattern (MBP), which describes motion information in the spatio-temporal space. An MBP is inspired by a Volume Local Binary Pattern (VLBP) and also computed from three consecutive frames. By introducing MBPs we mainly address the problem of decreasing the amount of ambiguities of VLBPs and propose a novel method to describe motions.

In the second part, we evaluate Volume Local Binary Patterns regarding different computation strategies. VLBPs are a famous feature for describing characteristics in the spatio-temporal domain and widely used in the computer vision community [23,31,38,41]. VLBPs are easy to implement and efficient in their computation.

Contribution: The main novelty of the paper can be summarized as follows:

- Propose a novel feature type for action recognition.
- Incorporate a temporal sliding window.
- Provide a summary and evaluation about computation strategies.

The proposed approaches are evaluated on the single-view KTH dataset [29] and Weizman dataset [5,9] as well as on the IXMAS dataset [36,35] for multi-view action recognition. Figs. 1–3 show some example images of the datasets.

The paper is structured as follows: Section 2 gives an overview about related work and concludes state-of-the-art methods. Section 3 briefly describes the machine learning algorithm Random Forest. Motion Binary Patterns and Volume Local Binary Patterns are explained in detail in Sections 4 and 5 while several computation strategies are proposed in Section 6. Section 7

☆This work has been partially funded by the ERC within the starting grant Dynamic MinVIP (Grant no: 277729).

* Corresponding author.
E-mail address: baumann@tnt.uni-hannover.de (F. Baumann).

http://dx.doi.org/10.1016/j.neucom.2015.03.097
0925-2312/

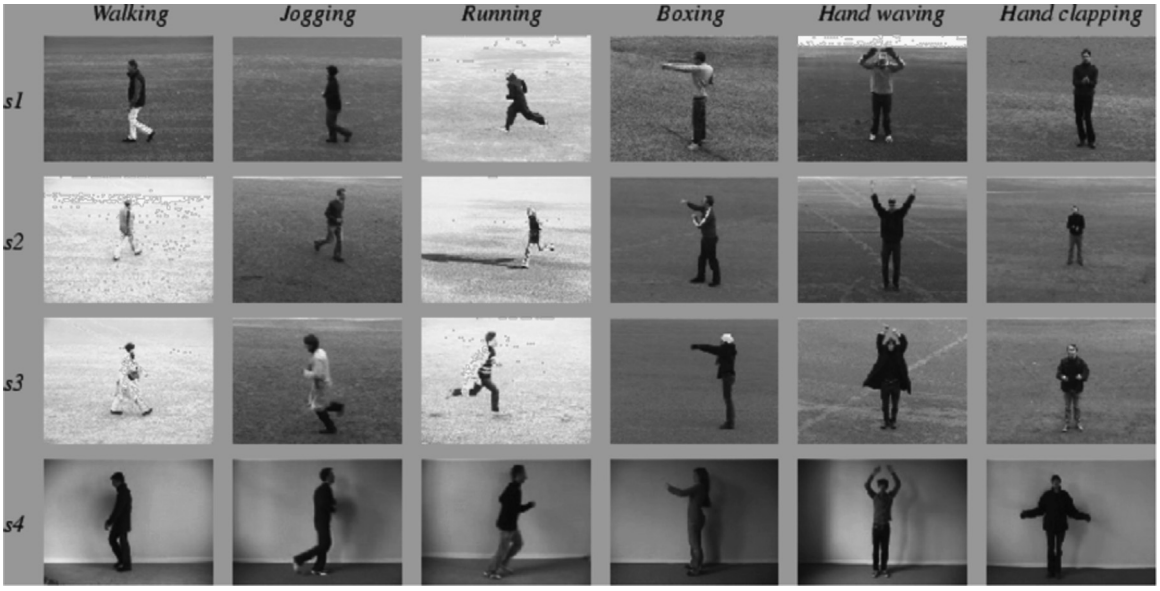

Fig. 1. Example images of the single-view KTH dataset [29]. The dataset contains six actions performed by 25 people under different conditions.

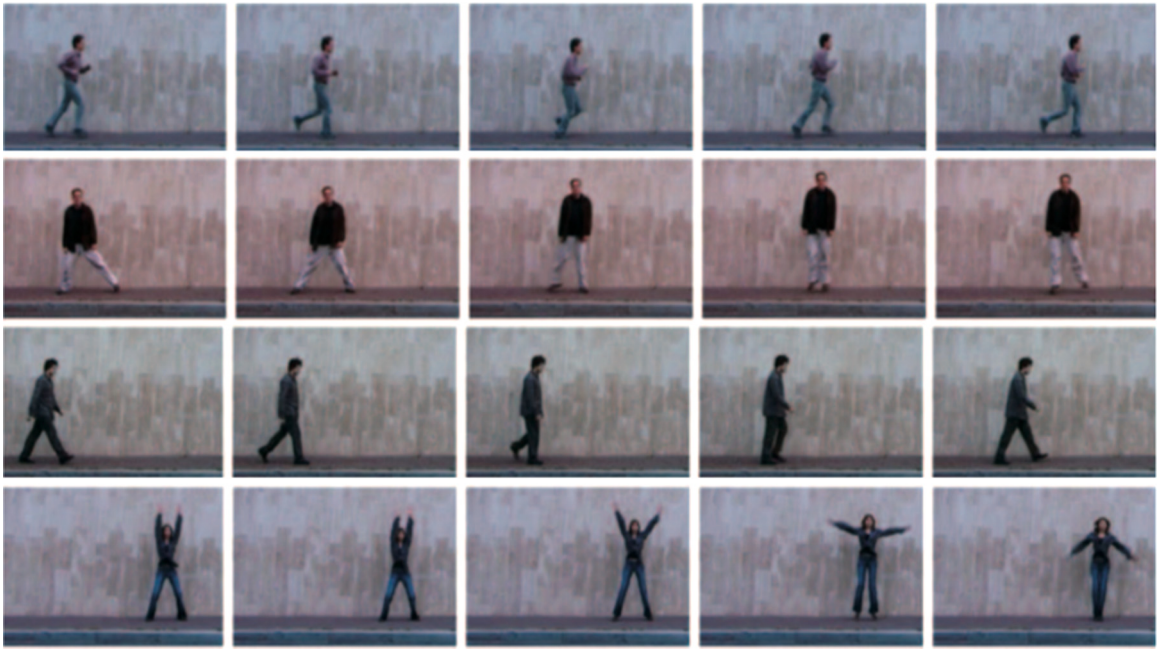

Fig. 2. Example images of the single-view Weizman dataset [5,9]. The dataset contains nine actions performed by nine people.

presents the experimental results and Section 8 concludes the paper and reveals some future work.

This paper extends the earlier published conference articles presented in [3,4]. Additionally, a more detailed introduction and more experiments for the proposed features are given.

2. Related work

Volume Local Binary Patterns (VLBPs) were proposed by Zhao and Pietkäinen to recognize dynamic textures [41]. VLBPs are an extension of the LBP operator that is widely used for several tasks of computer vision. A VLBP uses three parallel planes where the surrounding pixels are compared to the center pixel of the middle one. A VLBP takes the co-occurrences of all neighboring pixels into account leading to a huge feature vector and to ambiguities. Zhao and Pietkäinen additionally propose the LBP-TOP operator as an extension of the VLBP to avoid a huge feature vector and to decrease ambiguities [40]. Experiments were conducted with an application of recognizing facial expressions. Mattivi and Shao applied LBP-TOP [23,31] to the task of human action recognition. The authors reached 88.19% accuracy on the KTH dataset and by combining Extended Gradient LBP-TOP with Principal Component Analysis in [23] they reached 91.25%. Best results (92.69%) were achieved by combining Extended Gradient LBP-TOP with Dollar's detection method [31]. In this paper, we decided to use VLBPs instead of LBP-TOP for a comparison to the proposed novel feature operator. LBT-TOP reduces the size of the feature but do not achieve competing accuracies.

Liu and Yuen [19] proposed a boosted EigenActions framework which calculates a spatio-temporal information saliency map