Neurocomputing 188 (2016) 71–81

Contents lists available at ScienceDirect

Neurocomputing

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

People recognition in multi-cameras using the visual color processing mechanism

XiaoJin Lin, QingXiang Wu *, Xuan Wang, ZhiQiang Zhuo, GongRong Zhang

Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, College of Photonic and Electronic Engineering, Fujian Normal University, Fujian, Fuzhou 350007, China

ARTICLE INFO

Article history:
Received 8 July 2014
Received in revised form
16 September 2014
Accepted 7 October 2014
Available online 12 December 2015

Keywords:
Visual color processing
Moving people recognition
Spiking neural networks
Support vector machine
Multi-camera video datasets

ABSTRACT

In this paper a people recognition algorithm is proposed, in which the color processing mechanism is inspired by the biological visual system. The algorithm is constructed in two parts. In the first part a spiking neural network is proposed to extract the color features of the people images which are captured from videos, and a set of new features is generated by fusing the color features and color moments. In the second part, after a feature reduction, a Support Vector Machine is trained and then used to recognize a specific people. The algorithm has been successfully applied to recognize people in CASIA Database with a high recognition rate. In order to evaluate performance and analyze characteristics of people recognition algorithms in multi-camera scenes, Multi-Camera Video (MCV) dataset is made in this paper. It is used to evaluate and analyze the proposed method and a set of characteristics of the proposed algorithms are obtained. Experimental results demonstrate that the algorithm is comparable to state-of-the-art approaches in terms of accuracy and the direction for further improvement of the proposed algorithm is provided.

1. Introduction

Human visual perception has inherent advantages in information processing and identification. Many visual tasks are very simple tasks for human such as object recognition and classification, edge detection, motion tracking etc., but they are complex for computers. As fast development of computer technology and enhance of computer image processing capabilities, a number of issues such as large-scale image classification and identification technology, especially human detection and face recognition technology, attract attentions of researchers in the world. For example, the task of person re-identification has been studied in [1]. Tao et al. presented regularized smoothing KISS metric learning (RSKISS) which have obtained very good performance for person re-identification on the VIPeR dataset [1]. They proposed a new scheme for human behavior recognition on Wireless sensor networks, in which WSNs deploy an utmost scalable, flexible, and robust human behavior recognition system [2]. Xu has studied the problem of detecting sudden pedestrian crossings to assist drivers in avoiding accidents with low false alarm rate and high speed [3]. Wang proposed a new approach of automatically transferring a generic pedestrian detector to a scene-specific detector in static video surveillance without manually labeling samples from the target scene [4]. Iwashita provides a demonstration of person identification based on gait recognition from shadow analysis, which describes compensation steps to address shadow variation and defines measures of shape variation [5]. Wren has studied the indoor single track real-time tracking system named as Pfinder using small regions characterized [6]. Several well-known research institutions (e.g. Carnegie Mellon University, David Sarnoff research centers) develop the video surveillance and monitoring system VSAM [7]. State Key Laboratory of Beijing Pattern Recognition Institute of Automation under the Chinese Academy of Sciences (CASIA) achieved some good results in visual monitoring of the moving people. Current portrait recognition methods mainly rely on face recognition, some latest face recognition works are shown in [8–11]. The highest face recognition rate is 97.13%, 96.78%, 98.10% and 96.45%. In real life, the actual resolutions of human face images are often inadequate for using the algorithms because of the camera insufficient, the long distances and the undesirable illumination. These adverse factors often result in difficulty for utilization of face recognition. Therefore, this paper focuses on identification of the overall portrait, and it makes important significance in tracking people's whereabouts between multiple cameras. We study the people recognition in details in two aspects which include color feature extraction by spiking neural network and recognition algorithm based on color features, the highest recognition rate reach at 99.34% in the test using the CASIA Database A. The CASIA Database is made from one camera

* Corresponding author.
E-mail addresses: xiaojin900912@163.com (X. Lin), qxwu@fjnu.edu.cn (Q. Wu).

http://dx.doi.org/10.1016/j.neucom.2014.10.107
0925-2312/

Fig. 1. The five scenes.

in a very ideal setting. Many recognition algorithms can achieve very high recognition rates in evaluations using the dataset. In real surveillance systems, there are multiple cameras. Police often captures a criminal person in a camera on a criminal site, and then is trying to identify the criminal person in the videos from other cameras. In order to simulate this situation, we make a pedestrian dataset from Multi-Camera Videos (called MCV dataset) which can be used to generate different training and test datasets for different evaluation schemes to test and evaluate people recognition algorithms. The different evaluation schemes can be generated to analyze advantages and disadvantages of people recognition algorithms. The proposed algorithm has been analyzed using this dataset. The experimental results have indicated advantages and disadvantages of the proposed algorithm. The guidance of improvements has been proposed.

Compared with other visual features, color feature has high robustness for less dependence on the size, orientation, viewing angle of the image itself. Spiking neural network (SNN) is applied to extract color feature in this paper. The SNN is often regarded as the third generation of artificial neural networks. In this paper a simplified conductance-based integrate-and-fire model [12] is used for each neuron in SNN. There are various receptive fields in the striate cortex, retina and lateral geniculate nucleus [13–14] in human visual system. Signals from photoreceptors are processed by retinal interneurons, integrated by retinal ganglion cells and sent to the brain by axons of retinal ganglion cells [15]. The visual images are transferred among these neurons in the form of spiking trains through the ON or OFF pathways [16]. An ON response stimulates the emission of spikes and an OFF response restrains this behavior. Generally, the ON/OFF type neurons can be used to construct specific ON/OFF pathways in the visual system. Different ON/OFF pathways are used to construct the specific networks for color feature extracting in a biological manner. Thus a moving people recognition algorithm based on the visual color processing mechanism is proposed in this paper and it obtains a good result in the mainstream pedestrian gait databases.

Compared with state-of-the-art in people recognition domain, new contributions are (1) This paper apply biological inspiration visual processing in moving people recognition process. This mechanism has not been studied in current recognition algorithms. (2) A new feature extraction method is proposed to integrate the SNN outputs and color moments so that the proposed algorithm takes the advantages of SNN in key feature extraction and the advantages of color moments in the statistics. (3) We have developed a MCV dataset which can be used to simulate multi-camera surveillance. A method of producing test dataset is proposed to generate multiple evaluation schemes for test of pedestrian recognition algorithms. As various videos in very poor light conditions and backgrounds have been collected in the dataset to generate evaluation schemes, recognition algorithms are difficult to achieve very high recognition rates so that the dataset can be used to analyze the advantages and disadvantages of recognition algorithms.

Table 1
The number of people in every scene.

	Scene1	Scene2	Scene3	Scene4	Scene5
Back appearance	9	17	12	6	4
Front appearance	15	17	9	9	14

Table 2
The frame number of each scene.

	Scene1	Scene2	Scene3	Scene4	Scene5	
Back appearance	992	1666	1663	559	336	Total: 5216
Front appearance	1033	1084	1204	608	1051	Total: 4980
Total: 10,196						

The remainder of this paper is organized as follows. In Section 2, we propose a Multiple Cameras Videos datasets named MCV dataset which are used to simulate multi-camera surveillance systems. Section 3 is some related works include people detection and image preprocessing, we segment the human body color image into 3 parts. In Section 4, the architecture of the spiking neural network is proposed and SNN is used to extract eight ON/OFF pathways color features for each part of people. Feature statistics are introduced in Section 5. Fusion and recognition algorithms are introduced in Section 6. Experimental results and analysis are presented in Section 7. Section 8 gives a conclusion and topics for further study.

2. CASIA Database and Multi-Camera Videos dataset

This paper adopts the CASIA Database A [17] to verify the algorithm and do comparison with some other algorithms. The CASIA Database A is composed of 20 people corresponding to three kinds of angles (0-angle, 45-angle, 90-angle), each angle includes four image sequences. In this paper, the 0-angle image sequences are used to verify the algorithm.

In order to test and verify people finding algorithms in a more realistic condition and investigate advantages and disadvantages of people finding algorithms, we have made a new pedestrian dataset in which videos are captured from a surveillance system with five cameras corresponding to five scenes, the pedestrian dataset is named MCV dataset. The videos in the MCV dataset contain complex lighting conditions such as sunshine, backlighting, etc. And the people walk far away from camera to near camera. The five scenes from the five cameras are shown in Fig. 1.

In the MCV dataset, 39 people appear in different videos. The number of people in every scene is shown in Table 1. After data cleaning, 10,196 frames in 128 videos are taken from the five cameras. Frames of each scene are shown in Table 2. The total number of frames is more than twice the CASIA Database A (0-angles).