Neurocomputing 273 (2018) 481–493

Contents lists available at ScienceDirect

Neurocomputing

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

Counter-propagation artificial neural network-based motion detection algorithm for static-camera surveillance scenarios

Bo-Hao Chen[a,b,*], Shih-Chia Huang[c,*], Jui-Yu Yen[c]

[a] Department of Computer Science and Engineering, Yuan Ze University, Taoyuan 320, Taiwan
[b] Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan 320, Taiwan
[c] Department of Electronic Engineering, National Taipei University of Technology, Taipei 106, Taiwan

ARTICLE INFO

Article history:
Received 19 May 2015
Revised 3 July 2017
Accepted 1 August 2017
Available online 24 August 2017

Communicated by Dr. X. Li

Keywords:
Motion detection
Video surveillance
Neural network

ABSTRACT

Motion detection plays an important role in most static-camera video surveillance systems, yet video communications over wireless networks can easily suffer from network congestion or unstable bandwidth, especially for embedded applications. A rate control scheme produces variable bit rate video streams to match the available network bandwidth. However, effectively detecting moving objects in a variable bit rate video stream is a considerable challenge. This paper proposes an advanced approach based on a counter-propagation artificial neural network to achieve effective moving-object detection in such conditions. Qualitative and quantitative tests over real-world limited bandwidth networks show that the proposed method substantially outperforms other state-of-the-art methods.

1. Introduction

Video surveillance systems, which are usually implemented with static cameras, have a wide range of scientific and technological applications, such as computer vision [1], elderly care [2], traffic monitoring [3], and many others [4]. Video surveillance system applications include motion detection [5], identification [6], object classification [7], behavior recognition [8], and activity analysis [9]. Motion detection, in which moving objects are extracted from video streams, is an essential process in video surveillance systems and many approaches have been proposed to achieve complete and accurate detection in sequences of normal visual quality.

There are three major categories of conventional motion-detection approaches: optical flow, temporal difference, and background subtraction. Optical flow can achieve detection by projecting motion onto the image plane with proper approximation. However, this method is highly sensitive to noise and can be computationally inefficient. Temporal differencing detects moving objects by calculating the difference between consecutive frames, but it has a tendency to extract incomplete shapes of these moving objects, particularly when they objects are motionless or exhibit limited mobility in the scene. Background subtraction detects moving objects in a video sequence by evaluating the pixel feature differences between the current image and a reference background image. Of the various motion-detection strategies, background subtraction is the most widely used for the effective extraction of motion information from streaming video [10].

The need for precise motion detection has increased dramatically since the 9/11 attacks, which has subsequently led to higher demand for a more reliable and accurate background model generated through background subtraction. Consequently, many background-subtraction-based methods have been proposed to segment moving objects in a video sequence. A sigma difference estimation (SDE) approach to background subtraction, which employs a $\Sigma - \Delta$ filter technique, was proposed in a previous study [11]. The SDE method is used to calculate two orders of temporal statistics for each incoming pixel, providing a pixel-level decision model. However, often only a single $\Sigma - \Delta$ filter for each incoming pixel is used to model the reference background in the detection of moving objects, which can be insufficient when dealing with certain complex scenes, such as those featuring many moving objects. Therefore, the multiple SDE (MSDE) approach was proposed to detect moving objects in complex scenes by combining $\Sigma - \Delta$ estimates with different frequencies or phases [12]. Of the methods mentioned thus far, one of the most popular for background modeling is the Gaussian mixtures model (GMM) [13]. When GMM is used, each pixel value is modeled independently; the method then describes the distribution of each pixel that belongs to the background pixel category. Jodoin et al. proposed the simple statistical difference (SSD) method to accomplish motion

* Corresponding authors at: Department of Computer Science and Engineering, Yuan Ze University, Taoyuan 320, Taiwan, and Department of Electronic Engineering, National Taipei University of Technology, Taipei 106, Taiwan.

E-mail addresses: bhchen@saturn.yzu.edu.tw (B.-H. Chen), schuang@ntut.edu.tw (S.-C. Huang), tim904021@hotmail.com (J.-Y. Yen).

http://dx.doi.org/10.1016/j.neucom.2017.08.002
0925-2312/

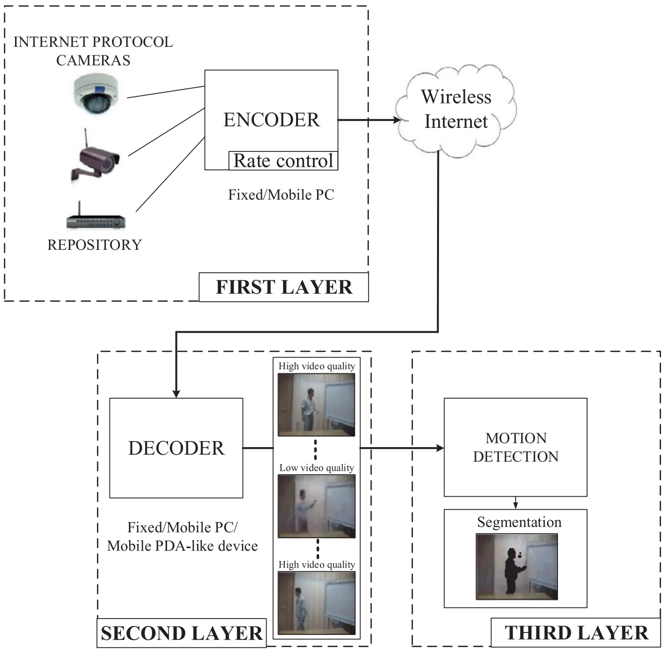

Fig. 1. Video surveillance systems operating in many real-world limited bandwidth networks.

detection by using the mean value and standard deviation [14]. The SSD method computes the mean value with standard deviation for individual pixels of the incoming frame to determine whether each incoming pixel should be labeled as belonging to either the background or a moving object. The multiple temporal difference (MTD) approach for motion detection maintains various previous reference images to reduce fractures inside moving entities [15]. The MTD method can effectively cope with the fractures and shadow regions of moving objects.

A recent method (ViBe) proposed by Barnich and Van Droogenbroeck [16] builds its background model from each pixel by randomly selecting color samples from the pixel's neighborhood and further correcting the model through a stochastic update strategy as frames advance. St-Charles et al. developed the SuBSENSE method [17], which is based on the pixel-based adaptive segmenter and uses color and local binary similarity patterns to separate moving objects from backgrounds. In another recent method called LOBSTER [18], the local binary patterns operator was employed and improved to resist shadow disturbances. To this end, LOBSTER adjusts the threshold to discriminate between the grey values of a center pixel and its neighboring pixels. In general, the background models produced by ViBe, SuBSENSE, and LOBSTER can provide passable detection results for videos communicated over an ideal network bandwidth.

In recent years, research conducted into video surveillance systems has been oriented toward the low-quality video streams prevalent in many real-world limited bandwidth networks. As illustrated in Fig. 1, handheld media and mobile devices have gained popularity, as have real-time video applications on wireless networks such as video conferencing and security monitoring. However, video communications over wireless networks can easily suffer from network congestion or unstable bandwidth. The quality of network services is seriously degraded whenever the traffic exceeds the available amount of network bandwidth. Rate control is an important video coding tool that attempts to lower video quality and produce variable bit rate video streams to match the available wireless network bandwidth, thereby minimizing network congestion [19–21]. In general, most background subtraction methods suffice for situations involving normal video quality [22–24]. However, complete and accurate motion detection in variable bit rate (VBR) video streams is a considerable challenge because the frequently changing video quality may compromise the applicability of the generated background models [25]. For example, the ViBe method randomly maintains 20 background models (as suggested in [16]) to detect moving objects in such videos. Due to the use of the stochastic update strategy, most of these background models still employ information from previous video quality levels to detect moving objects when the video quality changes suddenly. This may result in the generation of either artifacts or noise (i.e., false positives) in the produced binary mask.

In response, we propose a novel motion-detection method based on a counter-propagation artificial neural network (CPN) to segment moving objects in VBR video streams. Our method can detect moving objects more precisely than other background subtraction methods can, because of its ability to adapt to VBR changes and achieve complete and accurate detection for both low- and high-quality video streams. The CPN possesses both a feature-mapping network and basic competitive network with minimal structure [26]. This permits the CPN to provide a close fit for the adaptive background model for VBR video streams by employing an unsupervised learning process. Moreover, an online learning process adjusts the weights quickly. This allows for suitable motion detection in both low- and high-quality video streams. According to our experiments, the proposed method can provide more efficient and precise detection compared to other state-of-the-art methods for real-world limited bandwidth networks in a wide range of natural video sequences.

The remainder of this paper is organized as follows. Our proposed motion-detection method, along with its framework, is described in Section 2. The experimental results of our method are compared with other motion-detection methods in Section 3. Finally, concluding remarks are presented in Section 4.

2. Proposed motion detection approach

In our application, the CPN architecture is considered an adaptive background model by which effective detection of moving objects can be achieved in both low- and high-quality video streams. In the following sections, we discuss the two major modules of our proposed CPN-based motion detection (CPNMD) approach: a various background generation (VBG) module for training processes and a moving object extraction (MOE) module for recall processes.

2.1. Various background generation

In general, CPN consists of the input, Kohonen, and Grossberg layers [26] (Fig. 2). The neurons of the input layer are connected to each neuron of the Kohonen layer. Since the CPN is an unsupervised winner-take-all competitive learning network, the input patterns are categorized by each neuron in the Kohonen layer according to the winner-take-all rule. The Grossberg layer produces the corresponding outputs for the category. According to these principles of CPN, the adaptive background model can be constructed from the time series of each incoming pixel in every frame to express the properties of various bit rate video streams.

For each incoming frame I_t, let $(p_t^Y, p_t^{C_b}, p_t^{C_r})$ respectively represent the luminance, the blue-difference chroma, and the red-difference chroma component values of a pixel $p_t(x, y)$ as the input patterns in the input layer of CPN. The winning neuron is selected as the one with the longest distance between the input layer and the Kohonen layer. For each incoming pixel $p_t(x, y)$ of the tth frame