Int. J. Electron. Commun. (AEÜ) 68 (2014) 227–236

Contents lists available at ScienceDirect

International Journal of Electronics and Communications (AEÜ)

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

Human recognition system for outdoor videos using Hidden Markov model

Ansuman Mahapatra*, Tusar Kanti Mishra, Pankaj K. Sa, Banshidhar Majhi

Department of Computer Science and Engineering, National Institute of Technology Rourkela 769008, Odisha, India

ARTICLE INFO

Article history:
Received 20 November 2012
Accepted 17 August 2013

Keywords:
Pattern recognition
Fuzzy inference system
Background subtraction
Hidden Markov model
Vector quantization

ABSTRACT

Human recognition is an essential requirement for human-centric surveillance, activity recognition, gait recognition etc. Inaccurate recognition of humans in such applications may leads to false alarm and unnecessary computation. In the proposed work a robust background modeling algorithm using fuzzy logic is used to detect foreground objects. Three distinct features are extracted from the contours of detected objects. An unique aggregated feature vector is formed using a fuzzy inference system by aggregating three feature vectors. To minimize computation in recognition using Hidden Markov model (HMM), the length of final feature vector is reduced using vector quantization. The proposed method is explained using five basic phases; background modeling and foreground object detection, features extraction, aggregated feature vector calculation, vector quantization, and recognition using Hidden Markov model.

1. Introduction

Surveillance has become a major concern for the law enforcement agencies since the beginning of this century. The world has seen and experienced a series of dreadful incidents in the last few decades. This has fueled the agencies to use better surveillance mechanisms. This need become a challenge for the computer vision community to build foolproof systems. Among the various surveillance mechanisms, visual surveillance is the most widely used. It involves use of a set of cameras and monitoring the feed manually or automatically.

Surveillance of outdoor videos have some inherent challenges. Illumination variation, crowd, various types of objects etc. are to name a few that compounds the problem. Outdoor videos are obtained from cameras that are generally deployed outside buildings, by roads, in park and public areas, in fields etc. It is not necessary that humans will only be visible in the outdoor videos. Cars, motor cycles, buses, trucks, animals, birds etc. are also part of the other objects that are visible in the out door videos. Any good algorithm can easily detect all foreground objects from the background. However, all foreground objects are not of interest when it comes to human-centric surveillance. Such system demands recognition of humans in the presence of various other objects. In this paper we propose to detect all foreground objects in a video sequence followed by recognizing who among them are humans.

In the recent past the scientific community have proposed quite a few number of works. Lalonde et al. [1] have used a commercial PTZ camera to automate object tracking and recognition task for some surveillance tasks. The module can discriminate between various moving objects and identify the presence of pedestrians or vehicles, track them, and zoom on them in near real-time. de Léon and Sucar [2] use Fourier descriptors for human silhouette recognition. They have made an analysis of which and how many descriptors are enough to have a general human silhouette representation. A reduced number of components is sufficient for representing a human silhouette for recognition in different posses. Zhang et al. [3] have used *k*-means clustering for automatic labeling and Gaussian assumption for classifying moving objects of traffic scene videos into vehicles, bicycles and pedestrians. The classifiers are initialized and refined on-line automatically, instead of supervised learning and manual labeling of large training samples. See et al. [4] have used a fuzzy rule-base classification based on detected blob regions for human motion detection. Three blob characteristics are extracted from the major blobs using motion estimation and ellipse fitting. These characteristics are fed into a fuzzy rule-base for classification of the detected motion. A real-time visual surveillance system for detecting and tracking multiple people was proposed by Haritaoglu et al. [5] for monitoring the activities in an outdoor environment. It employs shape analysis and tracking to locate the position of people and create appearance models. It can determine whether a foreground region contains multiple

* Corresponding author. Tel.: +91 8895590157.
E-mail address: helloansuman@gmail.com (A. Mahapatra).

1434-8411/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.aeue.2013.08.011

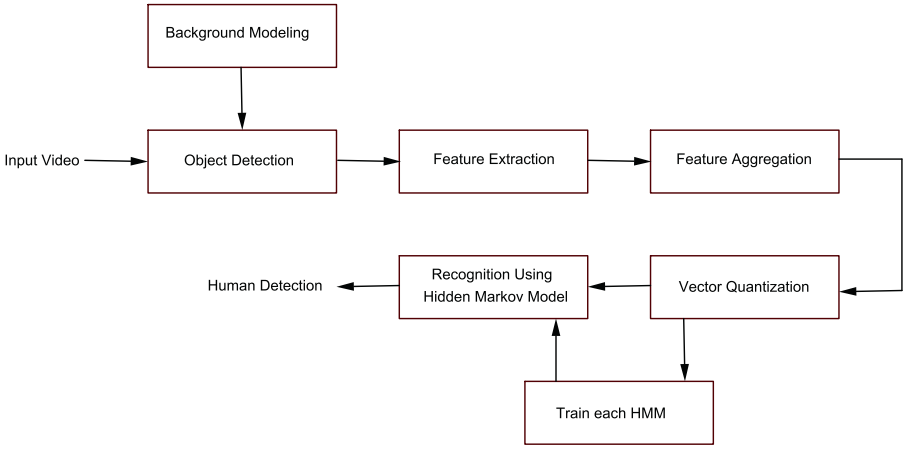

Fig. 1. An overview of proposed approach.

people. Kim and Uddin [6] proposed an activity recognition technique using LDA and PCA as feature and HMM is used for activity recognition.

Background subtraction technique has been widely used in foreground pixel detection where a fixed camera is usually used to observe dynamic scenes. Generating a background model from video sequences is a challenging task. A background estimation procedure is performed to separate motion from the background. Many standard methods exists for background modeling and segmentation of foreground objects [7–10]. In these methods, a background scene model is statistically learned using the redundancy of the pixel intensities in a training stage. The foreground objects can be detected if their gray values differs significantly from the background model.

Human detection in an image could be useful for human–robot interaction and human activity recognition. There is much related work carried out regarding human activity recognition using silhouette of a human and gait [11,12]. Human silhouette detection is based on motion, but usually it is not verified that the extracted silhouette belongs to a human. So some pattern recognition techniques can be embedded into the model so that it can identify, which is human contour and, which is not.

Significant works have been done for recognizing human activity and gait, but classifying human among the foregrounds still needs considerable attention. It is necessary to identify a human before identifying its activity. In this paper a simple and efficient method is proposed to recognize humans among other foreground objects. The proposed framework has five phases namely, (a) background modeling and object detection, (b) feature extraction, (c) feature aggregation, (d) vector quantization, and (e) recognition. Fig. 1 depicts the basic steps involved in human recognition. Using this framework, a smart surveillance system has been developed for outdoor surveillance videos.

The first part serves to detect all foreground objects in a video sequence after modeling the background. A new robust background modeling and object detection method is proposed using fuzzy inference system. In the second phase contour of the foreground objects are obtained from the silhouette. Then three types of features are extracted from the contour of the detected object. In the third phase, an aggregated feature vector is created from the three different feature vectors extracted in the previous stage using fuzzy inference system. In the fourth phase, the aggregated feature vector is converted into symbols using vector quantization to reduce the computational costs. In the final stage HMM for each class is learned from training samples and then test samples are given for recognition of human.

Rest of the paper is organized as follows. Section 2 presents the object detection algorithm. The feature extraction techniques are illustrated in Section 3. The fuzzy model for feature aggregation is presented in Section 4. Vector quantization for feature reduction is described in Section 5. The steps for recognizing human contours are outlined in Section 6. Simulation results are discussed in Section 7. Finally, Section 8 concludes the paper.

2. Background modeling and contour extraction

The silhouette of all foreground objects has to be extracted after detecting them by their motion in the surveillance area. It is assumed that video sequences are captured from a still camera. A background modeling technique is proposed using a fuzzy inference system. The existing object detection methods work on the gray-scale videos. But this causes informations loss while converting from color to gray. The main problem comes when background and foreground both have approximately same gray values. Then it is difficult for the algorithm to differentiate between foreground and background pixel. Sometimes two different colors such as dark blue and dark violet color when converted to gray scale, their gray values becomes very close to each other, and it cannot be differentiated. However, if color images are considered, it is easier to differentiate the foreground and background color. The proposed background modeling algorithm works directly on the color frames of the video. The main cause to take fuzzy inference system is that the brightness of the background changes rapidly as it is outdoor surveillance video. Therefore it is difficult to work in gray scale video frames. However, color of an object is less sensitive to brightness change.

To create a background model of the scene, "N" number of background frames $\{f_1^{bg}, f_2^{bg}, \ldots, f_N^{bg}\}$ are used. The minimum value of $N = 1$ and more statistical property of the scene can be learned with increase in the value of N. Value of N depends on the dynamic nature of the scene. If the scene is less dynamic then the value of N may be less, if it is more dynamic, then more number of background frames are required for effective learning of the scene. The background modeling of the scene using fuzzy inference system is a simple and efficient method. The fuzzy inference system takes red, green, blue value of each pixel for every background frame to create the background model. The detail of the system is provided in Fig. 2.