Neurocomputing 100 (2013) 19–30

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

HoGG: Gabor and HoG-based human detection for surveillance in non-controlled environments

Cristina Conde, Daniela Moctezuma *, Isaac Martín De Diego, Enrique Cabello

Face Recognition and Artificial Vision Division, Universidad Rey Juan Carlos, C. Tulipán, S/N, 28934 Móstoles, Spain

ARTICLE INFO

Available online 1 June 2012

Keywords:
Human detection
Video surveillance
Gabor filters
HoG

ABSTRACT

A new method (HoGG) for human detection based on Gabor filters and Histograms of Oriented Gradients is presented in this paper. The effect of Gabor preprocessing is analyzed in detail, in particular the improvement experienced by the image's information and the influence exerted over the extracted feature. To compare the performance of the proposed method, several alternative algorithms for human detection have been considered. In order to evaluate these techniques in non-controlled environments, a collection of standard databases, well known in the surveillance research community, has been used: PETS 2006, PETS 2007, PETS 2009 and CAVIAR. An exhaustive test design has been built based on two complementary evaluations: an evaluation oriented to counting people and a novel evaluation oriented to identification. Moreover, with the purpose of studying the performance of the Gabor-based preprocessing, a test adding Gabor filters to other local feature extraction methods, such as Steerable filters and the SIFT method, has been implemented. The HoGG method has achieved a good performance regardless of the difficulty of the images (occlusions, overlapping, carrying baggage, etc.). The proposed method has surpassed the alternative techniques in most of the analyzed situations. When the Gabor preprocessing is introduced into other local feature extraction methods, they achieve a better detection of the relevant information by enhancing the human shape. The results show that using Gabor preprocessing in techniques based on features like orientation or magnitude of gradient improve their performance. Given the excellent results obtained by HoGG at the identification-oriented evaluation, the method presented in this paper should be taken into account in the future design of intelligent surveillance systems.

1. Introduction

As an active research topic in computer vision, visual surveillance in dynamic scenes attempts to detect, recognize, and track certain objects from image sequences, and more generally to understand and describe object behavior. The aim behind such topic is to develop intelligent visual surveillance in order to replace the traditional passive video surveillance that has proven ineffective when the number of cameras exceeds the capability of human operators to monitor them [13]. Since humans are the main actors in daily activities of interest, one of the main tasks in video surveillance systems is people detection. This task can be very complex in situations involving critical infrastructures with high density of people, like airports, subway and train stations. That is, places where there are many people and where any attack or suspect situations could turn out to be dangerous. Consequently, this problem has been an active area of research in recent years [38]. Visual surveillance is a challenging scientific problem and an important field of application for computer vision. With increasing processor power, more attention has been given to the development of real-time smart surveillance systems. In addition, surveillance cameras have already been installed in many locations such as highways, streets, stores, ATM machines, homes and offices, and are socially accepted. The ability to analyze and understand human motion and recognition of human activities is key for a machine to interact intelligently and effortlessly with humans in a social environment [32]. Given the recent growth in security needs due to terrorist attacks, thefts, etc., it is crucial to have a reliable system for detecting people. However, this is a challenging problem due to environmental situations like illumination changes, occlusions and the fact that human bodies are non-rigid and highly articulated [36]. In addition, working in non-controlled environments increases the complexity of human detection tasks. In fact, visual surveillance in dynamic scenes, especially of humans and vehicles, is currently one of the most active research topics in computer vision [13]. It has a wide spectrum of promising applications, including access control in special areas, human identification at distance, crowd flux statistics

* Corresponding author. Tel.: +52 689787797.
E-mail addresses: cristina.conde@urjc.es (C. Conde), daniela.moctezuma@urjc.es, dmocteo@gmail.com (D. Moctezuma), isaac.martin@urjc.es (I. Martín De Diego), enrique.cabello@urjc.es (E. Cabello).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2011.12.037

and congestion analysis, detection of anomalous behavior, interactive surveillance use of multiple cameras, etc.

Currently, several human detection algorithms display a good performance in controlled conditions (see for instance [21,5,39]). However, when these algorithms are applied to real scenarios there is a sharp decrease in their performance. Hence, the task of human detection in non-controlled environments remains unsolved [30].

There are different approaches to computer vision-based video surveillance systems. One common solution is to extract moving objects and then apply an object tracking technique, whilst another widespread approximation is more tailored to object recognition. The former solution may obtain a good performance in controlled environments, but in the case of non-controlled environments, where the kind of the detected object is more relevant, the latter approach may be more suitable. The approach proposed by this paper is focused on human recognition without tracking. Usually, object recognition systems have three different stages: background segmentation, features extraction, and classification. In the first step, background segmentation and movement detection are applied in order to extract the moving objects from an image. The effectiveness of this task has a big influence on the final result of the surveillance system. In the second step, the feature extraction of moving objects is performed [11]. There are different ways to carry out this task: by using extraction windows, human templates (full body or body parts), histogram features, etc. Finally, in the third step, the features are used to determine whether an object is or not a person. To perform this task, several classifiers have been used: AdaBoost, Support Vector Machines (SVM) and K-Nearest Neighbors algorithm, among others, all well known in the literature. In all computer vision systems the database selected to evaluate the system performance is highly relevant. Usually just one database is considered and the system parameters are fixed to the specific conditions of this database. That is, the generalization capacity of the system is not considered.

As main contribution of this paper, a new method for human detection in non-controlled environments is presented. The proposed technique is based on Histogram of Oriented Gradient (HoG) [5] and Gabor filters [20]. For simplification purposes, this new method will henceforth be called HoGG. The effect of Gabor preprocessing is analyzed in detail, in particular the improvement experienced by the image's information and the influence exerted over the extracted feature. In order to compare HoGG with other state-of-the-art human detection methods [5,21,35] from a global point of view, several standard databases have been considered. Thus, all methods are tested in a wide variety of real conditions. To study the performance of every human detection method in each database, a detailed experimental design has been put in place. Also, two evaluations have been carried out to compare the different methods. Firstly, an evaluation oriented to counting people, in addition to which a novel evaluation oriented to identification is proposed. Moreover, an evaluation of the effectiveness of preprocessing images with Gabor filters in some alternative methods, such as Steerable filters [9] and SIFT [25], is also presented. The paper is organized as follows. Section 2 presents a brief review of state-of-the-art human detection algorithms. Section 3 describes the databases used for the tests. The proposed method, HoGG, is detailed in Section 4. The alternative methods and the experiments are shown in Section 5. Section 6 shows the results and, finally, the conclusions are presented in Section 7.

2. Overview of human detection

Several works dealing with human detection have been published in the last few years. These works can be classified into two different categories: those that use feature extraction windows and those that use human body models, templates or any other human-figure pattern [30].

First, the window-based feature extraction methods are discussed. In [35], the well-known Viola & Jones method (V&J) is presented. Although this is a robust and extremely rapid object detection method focused on face detection, it has been commonly used for human detection tasks too. An alternative method for object detection that combines AdaBoost learning with local histogram features has been proposed [21]. The basic idea in the Histograms of Oriented Gradients (HoG) algorithm [5] is that the local object appearance and shape can be characterized rather well by the distribution of local intensity gradients or edge directions. Usually, HoG has been used as baseline in order to develop new object detection algorithms.

The work presented in [23] shows a learning-based sliding window-style approach. The authors propose a main approach to learning and classifying human/non-human image patterns by simultaneously segmenting human shapes and poses, and extracting articulation insensitive features. The shapes and poses are segmented by a probabilistic hierarchical part-template matching algorithm. The method proposed in [39] integrates the cascade-of-rejectors approach with HoG features to achieve a fast and accurate human detection system. These features are obtained by HoGs of variable-size windows that automatically capture salient features of humans. A method that combines HoG and features extracted from co-occurrence matrices is presented in [33]. The fusion of two types of features (a quantized vocabulary of spatio-temporal volumes and a quantized vocabulary of spin-images) is proposed in [24] for human detection and action recognition tasks. Base detectors and segmentors are designed with edgelet features as a central component, so that only the shape information is used. In order to determine the best features for detecting people, an evaluation of human detection features is carried out in [34]. The human detection method proposed in the present paper, HoGG (detailed in Section 4), belongs in the window-based feature extraction methods. The main innovation introduced by HoGG over the alternative methods is the inclusion of a Gabor based preprocessing step. This preprocessing helps to emphasize the human body shape and improves the posterior gradient accumulation done by the HoG algorithm.

Next, the algorithms that use human body models, templates or any other human figure pattern are briefly reviewed. One simple approach to this is the binary shape model presented in [3]. In this work, an upper body shape is matched to an edge modulus image by simple correlation after symmetry-based segmentation. A more sophisticated approach is the Chamfer system proposed by Gavrila et al. [10]. This algorithm arranges human body models or templates in a hierarchical structure where the similarity between two templates is defined by the Chamfer distance. To detect and classify moving objects, two alternatives have been presented in [13]: shape-based silhouettes and motion based on periodic human movement. In the shape-based classification, different descriptors of shape information such as points, boxes and silhouettes are proposed to classify objects like humans. In the motion-based classification, the periodic human motion serves as a strong signal to classify moving objects. A part-based human body representation is introduced in [36]. It uses edgelet features, a kind of silhouette-oriented feature. This approach is tolerant to partial occlusions, point of view changes and pose variations. A new two-stage human detection method involving weighted matching and verification is presented in [31]. Given a set of templates that describe a number of human postures, a detection window is used to find the best matching description of the image. The best matching template is the one with the shortest distance to the