Pattern Recognition Letters 000 (2016) 1–10

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis☆

Djamal Merad[a,*], Kheir-Eddine Aziz[b], Rabah Iguernaissi[a], Bernard Fertil[a], Pierre Drap[a]

[a] Aix-Marseille University, LSIS - UMR CNRS 7296, 163 Avenue of Luminy, Cedex 9, Marseille 13288, France
[b] Toulon University, SeaTech School, Avenue de lUniversite - BP 20132, La Garde 83957, France

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Multiple-people tracking
Head detection
Occlusion handling
Re-identification

ABSTRACT

Multiple objects (targets) tracking plays an important role in computer vision. It is considered as the first step in many artificial intelligence applications that are developed to analyze people behavior for either security or statistical purposes. The most important challenge faced by algorithms designed for multiple objects tracking is the identity switches that occur between tracked objects due to occlusions and interactions between these same objects. This work falls within the scope of video-based behavioral marketing analysis and aims to better understand the purchasing behavior of customers by analyzing their movements in a densely-populated sales area. We propose to use a re-identification strategy to prevent these identity switches. This re-identification strategy is based on segmenting detected individuals into head, torso, and legs in addition to the classification of their appearances into front and back poses. This re-identification module is integrated within our tracking system to fuse tracklets obtained from a particle filter based tracking framework in a mono-camera tracking system. The combination of these tracking and re-identification modules allows the recovery of global trajectories for tracked individuals.

1. Introduction

The understanding of customers' needs and the development of personalized offerings are perpetual issues for shopping centers and other commercial areas in order to stimulate purchases. In fact, seven out of ten purchase decisions are made inside the store itself. Marketing studies conducted by well-known agencies in the field show that purchases are in general related to the customers' behavior and habits such as their trajectory and gaze in the sales area. Thus, the study of customers' behavior and habits is becoming increasingly necessary when developing marketing strategies for distribution retailers.

Some major retailers are already using cameras to analyze the attendance of their store by customers at different times over the whole day to better adapt the number of employees required to be present in the sales area. In case, more details about customers' behavior are needed, human operators are sometimes asked to interpret recorded videos or directly observe customers' behavior in the store.

Therefore, new solutions for behavior analysis based on artificial vision are emerging. Among these solutions, we may mention gaze tracking ([22]) and gesture recognition ([27]). There is also a number of applications in which behavioral analysis is based on tracking people [12,14], and [36].

Pedestrians tracking in video stream are defined as the ability of localizing tracked individuals in each frame of an input video in order to record their trajectories. The problem of objects tracking in a mono-camera tracking system is approached either as a single object or multiple objects tracking. The single object tracking is focused on resolving the problems of scale and illumination variations, whereas the multiple objects tracking algorithms are focused on keeping the identities of tracked objects. The most important challenge that is faced in the field of multiple objects tracking is the identity switches that occur between tracked targets due to occlusions and interaction between tracked objects.

Many researches had been conducted in the field of multiple objects tracking to solve the problem of identity switches. Among these works, we may mention those focused in resolving severe occlusion such as the ones proposed in [5,8,40], and [17]. Our work lies in this category of approaches. In this work, we used our re-identification method proposed in [4] for people re-identification across non-overlapping camera views to prevent identity switches that are caused by long term full occlusions in a mono-camera tracking system.

☆ This paper has been recommended for acceptance by Cosimo Distante.

* Corresponding author. Tel.: +33 4 91 17 79 14; fax: +33 4 91 17 79 22.
E-mail address: djamal.merad@univ-amu.fr, djamal.merad@gmail.com (D. Merad).

http://dx.doi.org/10.1016/j.patrec.2016.04.011
0167-8655/

This paper is organized in three main sections. The first section is dedicated to an overview of the existing methods in the field of multiple objects tracking. In the second section, we describe our proposed strategy that is used to solve the problem. Whereas, the third and last section is dedicated to show some results of our tracking system.

2. Related works

In recent years, multiple objects tracking became one of the most active research areas in computer vision. Thus, several methods had been proposed to deal with different issues faced in this research field. The existing methods for multiple objects tracking may be classified into two categories according to the number of tracked objects. In this classification which is adopted by Luo et al. [23], we can distinguish the association based tracking (ABT) [1,7,26], and [19] in which the number of tracked objects is varying over time and the category free tracking(CFT) [16,45,46], and [44] in which the number of tracked objects is known for each frame of the video.

The association based tracking algorithms are proposed for tracking a specific type of objects such as pedestrians. In general, these methods use a detector that is able to detect the specific type of objects in video frames. Several methods had been proposed for this category. Among these methods we may mention those based on optimization such as the method proposed by Berclaz et al. [7] in which the problem of multiple objects tracking is solved by k-shortest paths optimization, and the methods proposed by Andriyenko and Schindler [1] and Andriyenko et al. [2] which are based on the minimization of energy functions. Berclaz et al. [7] used an object detector that produces the occupancy map for the positions occupied by all tracked objects in all frames of the video. Then, they introduce a generic multiple objects tracking framework in which the false positive detections are filtered and the gaps caused by missing detections are filled before using the k-shortest paths algorithm in the reconstruction of trajectories for tacked objects. Whereas, Andriyenko et al. formulated the tracking as the minimization of a unified discrete-continuous energy function. They joined the data association problem with trajectory estimation by minimizing a consistent discrete-continuous energy. Their algorithm is able to track varying number of people due to the fact that it is based on HOG detector which is used to detect individuals. Another approach used for this category of methods is the formulation of tracking as inference in conditional random field (CRF). For this approach Yang and Nevatia [43] proposed an online learning approach for multi-target tracking which is formulated as inference in conditional random field model. They created two models one is used to distinguish targets, and the other one is used to distinguish pairs of similar targets. The use of these two models allows the association of tracklets. In the same context, Milan et al. [26] proposed a mixed discrete-continuous conditional random field in which they took advantage of exclusion constraints which suggest that a detection cannot be assigned to two trajectories and that a trajectory cannot have more than one observation per frame. In the same category of methods Leal-Taixé et al. [19] proposed a method for multiple people tracking based on a model that captures interaction between peoples. For this, they introduced a string that encodes interaction between individuals from image features. This model allows encoding the effect of undetected individuals in the scene.

For the second category of approaches which is the category free tracking, the tracked objects should be initialized in the first frame of the video. These approaches are not designed for a specific type of objects. In this category, Zhang and van der Maaten [45,46] exploited the spatial constraints in model-free tracker by learning classifiers for tracked objects and structural constraints using structural SVM. They also proposed a new approach that may be used to improve the generic objects detectors that are used to track specific objects. Yang et al.[44] proposed a method in which the problem of multiple objects tracking is viewed as a game competition between different object trackers where each tracker tries to maximize its visual evidence and generates interferences to other trackers. Then, the problem is solved by finding the Nash equilibrium of the game. Whereas Shi and Karl [37] developed a real time level set implementation for multiple objects tracking. This region-based model uses tracking results from previous frames to initialize the curve and used an energy minimization function to locate tracked objects in the current frame. For the same category, Hu et al. [16] proposed an algorithm for multiple objects tracking with occlusions reasoning. In this method, the tracked objects are divided into several blocks for which low dimensional log-Euclidean Riemannian subspace models are learned. These blocks are used to update appearances of tracked objects such that just apparent blocks are updated. This strategy allows update of appearances for partially occluded objects.

Due to the fact that occlusions are one of the most important challenges faced in tracking multiple objects, several methods focused on solving this issue were proposed [5,8,40], and [17]. Bae and Yoon [5] proposed an online multiple objects tracking method. Their work focused on solving the problems of severe occlusions and similar appearances. In order to solve the problem of severe occlusions they developed a method based on tracklets confidence which is associate to an online discriminative appearance learning module which handles similar appearances in order to associate tracklets that belong to the same object. Chen et al. [8] proposed a method for tracking people in crowded scene with partial long term occlusions. Their algorithm is based on creating constrained sequential labeling that is used to label the mid-level video features with the tracked objects identifiers by respecting some hard constraints. In the same way, Izadinia et al. [17] developed a method which is based on simultaneous tracking of pedestrians and their body parts. In this work, the tracking is done in three steps. It starts by associating detected pedestrians to obtain tracklets and then perform association between the different parts of pedestrians in order to revert the pedestrians associations that do not conform with part associations. At the end, simultaneous association between all tracklets is performed. In the same context, for occlusions management Tang et al. [39,40] exploited the fact that most occlusions in crowded scenes occurs between two peoples to develop a method for tracking multiple people in crowded scenes with several occlusions. They created a double-person detector that is able to detect two persons who are occluding each other. They integrated this joint detector into the tracking strategies proposed by Andriyenko and Schindler [1] and Pirsiavash et al. [33].

The problem of re-identifying individuals after being lost in a tracking system is a very common problem which was treated in mono-camera system to recover identities of tracked individuals after long term full occlusions as we said previously. But also to associate tracked individual in an inter-camera context for multi-camera tracking with non-overlapping camera views. For this context, many research had been conducted and many methods were proposed. Among these methods, we may mention the learning based methods such as those using SVMs [3,29], and those using Adaboost [6], [18]. We may also mention the direct methods that rely on finding the most discriminative signature such as [6] for the texture signature, [15,41] for the shape, and [11,24] in which simple features (HSV and RGB histograms) are combined with the brightness transfer function (BTF) to improve the re-identification.

Even though many researches had been conducted and several methods had been proposed to solve the problem of identity switches caused by severe occlusions in multiple targets tracking

Please cite this article as: D. Merad et al., Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis, Pattern Recognition Letters (2016), http://dx.doi.org/10.1016/j.patrec.2016.04.011