Computer Vision and Image Understanding 144 (2016) 228–236

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

People re-identification across non-overlapping cameras using group features

Norimichi Ukita*, Yusuke Moriguchi, Norihiro Hagita

Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5, Takayama, Ikoma, Nara, 630-0192, Japan

ARTICLE INFO

Article history:
Received 16 December 2014
Accepted 26 June 2015

Keywords:
People re-identification
Non-overlapping cameras
People grouping

ABSTRACT

This paper proposes methods for people re-identification across non-overlapping cameras. We improve the robustness of re-identification by using additional group features acquired from the groups of people detected by each camera. People are grouped by discriminatively classifying the spatio-temporal features of their trajectories into those of grouped people and non-grouped people. Thereafter, three group features are obtained in each group and utilized with other general features of each person (e.g., color histogram, transit time between cameras, etc.) for people re-identification. Our experimental results have demonstrated improvements in people grouping and people re-identification when our proposed methods have been applied to a public dataset.

1. Introduction

Tracking via non-overlapping distributed cameras is crucial for efficient monitoring of people activities over a wide area. Such tracking can be achieved using re-identification methods across multiple cameras following visual tracking in each field of view.

In general, people re-identification in a surveillance scenario should be performed not by recognition using high-level image features, such as face recognition, but by low-level feature matching. This is because image regions of people captured by general surveillance cameras are typically too small for high-level recognition techniques to be applicable.

To improve people re-identification, we propose new features extracted from the group of people detected. The group of people (i.e., each rectangle shown in Fig. 1) is regarded as a set of people who are cohered based on the similarity of their relative positions and trajectories. Under the assumption that people in the same group are, in general, observed together even in different cameras,[1] the new features of each group are employed for people re-identification. To improve the accuracy of re-identification, our proposed method combines the proposed group features with traditional image cues that represent the appearance of each person.

The main contributions of this paper are summarized below.

- People grouping using spatio-temporal features of their trajectories: Our proposed spatio-temporal features and classification scheme achieve people grouping such that (i) the co-occurrence of different features are expressed and (ii) noisy and ambiguous features are removed.
- People re-identification using group features: The group features are represented by the trajectories of people in each group and the number of these people, as well as the image cues of people in the group.

2. Related work

For people grouping, their trajectories are used in the proposed method. These trajectories can be acquired by visually tracking people in a video. While people tracking in a dense crowd[2,3] has been increasingly important in computer vision, this paper focuses on scenes with a relatively-sparse number of people for detecting groups only from trajectory-based features.

For grouping people, the proposed feature is represented by spatio-temporal relationships between the trajectories of two pedestrians. The effectiveness of spatial relationships for grouping people has also been explored in still images [4,5]. Two key differences between using still images and videos are (1) temporal cues are available in videos but not in still images and (2) rich appearance features (e.g. age and gender estimation from a face) are available in still images, whereas it is difficult to extract such features from videos, because people are usually imaged smaller in videos.

As a model for representing interactions between people trajectories, the social force model [6], which adjusts a repulsive force between people depending on their social relationship, has been

* Corresponding author. Fax: +81 743 72 5279.
E-mail address: ukita@ieee.org, ukita@is.naist.jp (N. Ukita).

[1] This assumption is reasonable, for example, in indoor scenes where routes among cameras are limited (e.g. corridors in a building) and nearby cameras between which most pedestrians walk along similar trajectories.

http://dx.doi.org/10.1016/j.cviu.2015.06.011
1077-3142/

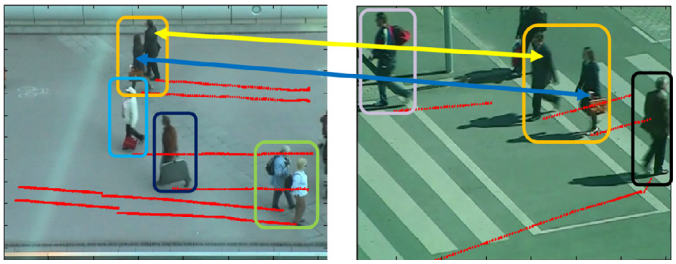

Fig. 1. Examples of images captured by surveillance cameras. For wide-area human tracking, identification across fields of view (indicated by solid arrows) is required, as well as tracking within each field of view (indicated by red dotted lines). Each rectangle represents a group. (For interpretation of the references to colour in this figures legend, the reader is referred to the web version of this article.)

widely used. This model is employed in several machine vision problems such as abnormal behavior detection [7] as well as people grouping [1,8].

In addition to models, a classification scheme is also crucial for people grouping. Bottom-up hierarchical clustering, iterative clustering using priors of collective behaviors of a group, and conditional random fields (CRFs) have been employed in [9], [14] and, [10], respectively. The classification scheme can also be formulated as a non-linear optimization problem for finding the optimal trajectories of groups [11,12]. These approaches are superior in terms of accuracy in group detection as compared with those using simple criteria (e.g. only proximity [13]).

For precise grouping in complex situations, the above-mentioned CRF-based method [10] optimizes its parameters by employing the annotation of people groups (i.e. a group to which each person belongs) in training data. These annotations reveal subtle differences of trajectories between people within and outside a group. With the annotation data, discriminative classification is also possible to improve the accuracy of people grouping [15].

For people re-identification, a number of image features have been proposed, for example, using a set of feature points [18] and histogram of oriented gradients (HOG) [20] based features [19]. Among all such approaches, features using a color histogram are robust against the change in viewpoint of a camera (e.g., a color histogram extracted from an HSV color space [21]).

In addition to the change in viewpoint, variations of illumination among cameras also make re-identification difficult. To cope with this problem, several learning techniques and metrics have been studied, including Adaboost [24], RankSVM [25], multi instance learning boosting[26], and distance learning[27]. While these techniques can improve re-identification across non-overlapping cameras, re-identification with poor appearance features in surveillance videos remains a difficult problem.

Unlike existing image features representing the appearance of people, our focus in this paper is on the features of a group of people. Appearance features of people in a group have been proposed in [33,34]; these features are extracted not only from a person of interest but also from people who are in the group with that person. Such features are based only on image features and have also been employed for visual tracking (e.g., in [35]). In addition to those image cues, our proposed features also consist of other properties of people in each group (i.e. the trajectories and the number of people in each group).

3. Overview

Fig. 2 illustrates the key processes of the proposed method. The entire process consists of two sub-processes, the process performed in each camera and the one that incorporates multiple cameras (these two sub-processes are enclosed by purple rectangles in the figure).

In an image obtained from each camera, several features of each pedestrian used for people re-identification are extracted. To obtain each pedestrian's region and trajectory, the pedestrian is visually detected and tracked.

While the algorithm of group detection has been proposed in [36], new experiments with image sequences are conducted for demonstrating its applicability to a surveillance camera scenario and the proposed people re-identification method.

For people re-identification across multiple cameras, conventional features extracted from the region of each pedestrian are also used, including color histograms (see Section 5.1.1) and transit times between cameras (see Section 5.1.2). In addition to these two conventional features, new group features are employed in the proposed

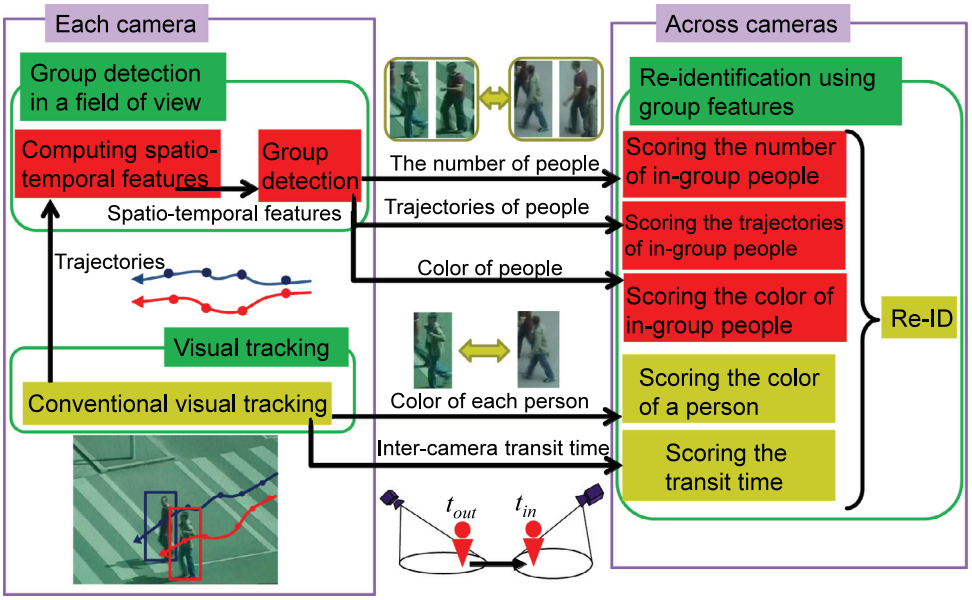

Fig. 2. Overview of the proposed people re-identification method. The proposed group features are computed in processes indicated by red rectangles. The three key components of the entire process are colored in green and are as follows: (1) visual tracking in each camera, (2) group detection in each camera, and (3) re-identification across cameras.(For interpretation of the references to colour in this figures legend, the reader is referred to the web version of this article).