AVSS 2011 demo session: Interactive Person-Retrieval in a Distributed Camera Network

Martin Bäuml[1]
[1]Institute for Anthropomatics
Karlsruhe Institute of Technology
Adenauerring 2, 76137 Karlsruhe, Germany
baeuml@kit.edu

Rainer Stiefelhagen[1,2]
[2]Fraunhofer Institute of Optronics,
System Technologies and Image Exploitation
Fraunhoferstr. 1, 76131 Karlsruhe, Germany
rainer.stiefelhagen@kit.edu

Abstract

Tracking and identifying persons in videos are important building blocks in many applications. For interactive investigation of surveillance footage it is often not even necessary to uniquely identify a person. It rather suffices to find occurrences of a person indicated by the user with an exemplary image sequence. We present a system in which the search for a specific person can be initiated by a sample image sequence and then be further refined by interactive feedback by the operator. The demonstrated system will track people online in multiple cameras and make the sequences immediately searchable from a central station.

1. Introduction

Video-based security systems monitor more areas than ever, generating vast amounts of surveillance footage, far too much to be looked at, searched and categorized by any human operator.

One possible use case in a surveillance setting is the search for occurrences of specific persons. Often it is not necessary to actually uniquely identify the person by name, but only find occurrences of the same person – as given by an example sequence – in other camera views and/or at an earlier time. For example, a human operator might want to know where a suspicious person has been during the last 10 minutes.

The general objective of our system is to find occurrences of specific persons in videos. We use facial appearance as feature in order to be independent of a person's clothing. This is important in the surveillance domain when the timespan of videos to be searched is longer than one day.

The person to be searched for does not need to be known and trained in in advance, but rather is specified by presenting the system with an example sequence of the person in question.

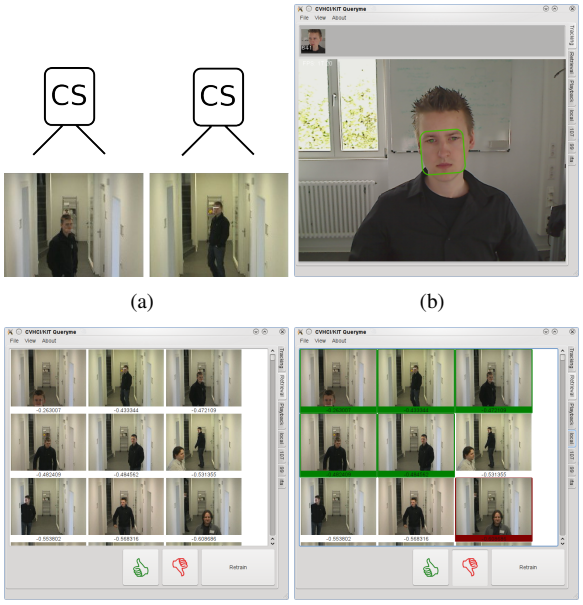

Figure 1. User interface of our retrieval system. (a) Persons are tracked by one or more cameras/capture stations. (b) For querying tracks of oneself, the user steps in front of the retrieval station. (c) The system reports and ranks the found sequences. (d) The user can give positive and negative feedback to improve the retrieval results.

2. System overview

The demo system consists of two capture stations and one retrieval station (cf. Figure 1). The capture stations are equipped each with a camera to track persons in real-time. These correspond to distributed surveillance cameras in a surveillance deployment. The retrieval station can be used by the user to find sequences of himself in the database, as captured before in the capture stations.

A user can start the query by stepping in front of the retrieval station. The retrieval station automatically tracks the

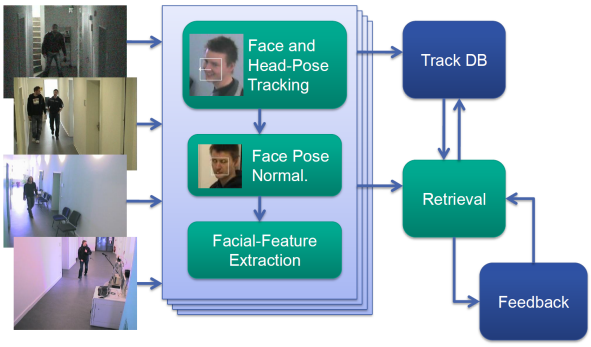

Figure 2. Overview over the retrieval approach.

person standing in front of it and accumulates the data necessary to search for other tracks of the same person. Finally, the user can refine the search by giving feedback on the results. Both positive and negative feedback is possible. The additional information is used to enhance the current query and then retrieve more accurate results.

3. Person retrieval in surveillance data

Person retrieval in surveillance data should work without time-expensive preprocessing of the data. After all, an operator might want to query the system to determine where a suspicious person has been during the last 5 minutes. Hence it is critical that (i) the tracking and feature extraction work in real time and (ii) first results to a query are available within a matter of seconds.

We use a given example sequence to train a support vector machine (SVM) specifically for each query. The SVM is then used to score each frame of each of the sequences in the database in order to find tracks close to the query track. This is very fast in practice (it takes less than 100 ms to search 1000 of possible target tracks with a linear SVM, and about 3 seconds with a polynomial SVM).

User feedback integrates well with this approach, as it can be used directly to enhance the positive and negative training set for the query classifier training.

For a more detailed description of the approach the reader is referred to [1].

4. Acknowledgments

This work has been funded in part by OSEO, French State agency for innovation as part of the Quaero Programme, and in part by the German Federal Ministry of Education and Research (BMBF) under project PaGeVi.

References

[1] M. Bäuml, K. Bernardin, M. Fischer, H. Ekenel, and R. Stiefelhagen. Multi-Pose Face Recognition for Person Retrieval in Camera Networks. In *International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, 2010.