



Doctoral Thesis

Visual surveillance Dynamic behavior analysis at multiple levels

Author(s):

Breitenstein, Michael D.

Publication Date:

2009

Permanent Link:

<https://doi.org/10.3929/ethz-a-005957048> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No. 18646

Visual Surveillance:
Dynamic Behavior Analysis
at Multiple Levels

A dissertation submitted to the
ETH ZURICH

for the degree of
Doctor of Science ETH

presented by
Michael D. Breitenstein
MSc ETH in Computer Science
born March 1st, 1980, citizen of Bettwil AG

accepted on the recommendation of
Prof. Dr. Luc Van Gool, ETH Zurich and K.U. Leuven, examiner
Dr. Ian Reid, University of Oxford, co-examiner
Prof. Dr. Bastian Leibe, RWTH Aachen University, co-examiner

2009

Abstract

New webcams and surveillance cameras are installed daily all around the world, adding huge quantities of data to the information stream that needs to be processed. As this happens, it is critical to develop methods that process such data-streams automatically and in real-time, reducing the manual effort that is still required for video content analysis. Of particular interest is to analyze the behavior of moving persons.

Different aspects of a scene can be recorded at different scale levels. While high-resolution portrait pictures facilitate the analysis of facial details, they do not allow us to observe the motion and interaction between different persons. In contrast, a camera observing an entire scene may result in images where individual persons are too small to be detected. However, spatial structure and motion patterns can still be extracted. Thus, each scale level contains useful information but requires a different approach to exploit it.

In this work, we present algorithms to analyze the behavior of dynamically moving persons in real-world scenes at different scale levels. The focus lies on developing methods that work online and fully automatically, adapting to the observations acquired so far.

We present five novel methods. First, we present a real-time method to detect unfamiliar faces and to estimate their poses. This algorithm is applicable to a large pose range and is robust to facial variations and expressions. The second innovation consists of a method that detects and tracks moving persons in complex scenes. The algorithm is suitable for a wide variety of online applications and does not require scene-specific knowledge. Another branch of this thesis introduces a statistical model to analyze the activity in a scene, where agents interact according to complex dependency patterns. It allows us to automatically find periodic actions and relations between co-occurring and consecutive activities. Then we present a technique to detect novelty from webcam footage at the scene level. During run-time the method automatically learns what usually happens in a scene and continuously adapts to the scene content. Finally, we propose an approach to estimate the scene structure from videos. This estimation may be used to guide an object detection system to those image regions and local scales at which objects are more likely to occur.

Zusammenfassung

Obwohl laufend neue Überwachungs- und Internetkameras installiert und deren Datenströme im Internet verfügbar gemacht werden, können diese nicht oder nur beschränkt analysiert werden, da der (manuelle) Aufwand zu gross ist. Je nach Auflösungsstufe der Bilder werden verschiedene Aspekte einer Szene aufgenommen. So sind zum Beispiel in hochaufgelösten Portraitaufnahmen Gesichtsdetails erkennbar. In Aufnahmen von mehreren Akteuren sind diese nicht mehr sichtbar, dafür können die Bewegungen und Interaktionen zwischen verschiedenen Akteuren beobachtet werden. Auf einer weiteren Stufe überblickt die Kamera eine noch grössere Szene, in welcher die einzelnen Akteure nur noch schwerlich auszumachen sind, wo jedoch die Struktur der Szene geschätzt und Aussagen über Bewegungsmuster getroffen werden können. Jede Auflösungsstufe enthält somit nützliche Informationen, deren Analyse jedoch unterschiedliche Ansätze erfordert.

In dieser Doktorarbeit präsentieren wir Algorithmen zur Analyse des Verhaltens von Personen auf verschiedenen Auflösungsstufen. Den Fokus richten wir auf Ansätze, die es erlauben, Daten automatisch und zur Laufzeit zu analysieren und sich an vorherige Beobachtungen zu adaptieren.

Wir präsentieren fünf Innovationen. Zuerst präsentieren wir eine Methode um Gesichter und deren Posen zu erkennen. Dann wird ein Algorithmus vorgestellt, der Personen detektiert und verfolgt, die sich dynamisch durch komplexe Szenarien bewegen. Dabei wird keinerlei Vorwissen über die Personen oder die Szene benötigt, und die Methode ist für verschiedenartige Anwendungen einsetzbar. Ein weiterer Teil der Dissertation befasst sich mit der statistischen Analyse von Aktivitäten in einer Szene. Ein Modell wird eingeführt, welches die Abhängigkeiten von gleichzeitigen und aufeinanderfolgenden Bewegungen von Akteuren analysiert. Ausserdem präsentieren wir eine Methode, die Anomalien in laufenden Aufnahmen von Internetkameras findet, ohne vorher ein Modell der Szene gelernt zu haben. Schliesslich wird ein Ansatz vorgestellt, um die Struktur einer Szene zu lernen und das Auffinden von Personen zu erleichtern.