Chapter 7

TRACKING GROUPS OF PEOPLE FOR VIDEO SURVEILLANCE

Frederic Cupillard, Francois Bremond and Monique Thonnat

INRIA Sophia Antipolis, ORION group
2004, route des Lucioles, BP93 - 06902 Sophia Antipolis Cedex, France

Abstract We propose in this paper a method for tracking groups of people in a metro scene to recognise abnormal behaviours such as violence or vandalism. After presenting the overall system, we describe the tracking algorithm for groups of people. Finally we present results illustrating the algorithm.

Keywords: Long Term Tracking, Human Group Model, Knowledge Based System

1. INTRODUCTION

In this paper, we present a method for tracking groups of people in a metro scene to recognise abnormal behaviours such as violence or vandalism [5]. Given a video sequence, our proposed algorithm is able to track real groups of people present in the scene. This algorithm is composed of three steps. First, from the current image of metro scene, a motion detector detects mobile objects in the scene, stores each of them in a moving region structure which is classified into different types, such as person, group or noise. Second, a tracking module detects and follows the real groups present in the scene by computing the trajectories of moving regions that can correspond to persons inside a real group. Third an interpretation module recognises the behaviour of the tracked groups. In this article, we focus on the second step, tracking groups of people. Contrary to traditional trackers [1, 2, 3, 6] that track each per-
son individually, our algorithm tracks globally all the persons belonging to the same group and tracks these persons with a time delay to keep coherent the structure of the group during the entire sequence. We illustrate this algorithm with video sequences taken for the European project ADVISOR (http://www-sop.inria.fr/orion/ADVISOR/default.html).

2. MOTION DETECTOR

The goal of this step is to detect mobile objects in the scene and classify them into moving regions with a label corresponding to their type, such as person. This task can be divided into three sub-tasks: detection of moving regions, extraction of features, classification of moving regions.

First for detection of moving regions, we compute a difference image between a current image and a reference image [4]. Because the camera is fixed, the reference image is a still image representing the scene without mobile objects (also called background image). Second a thresholding of this difference image is realized, then a filtering and connected component analysis create all the moving regions (represented by their bounding boxes) that correspond to the mobile objects. The reference image is updated during the process to take into account illumination changes.

The extraction of features step consists in the computation for each moving regions of 8 parameters: centre of gravity, position, height and width all four defined both in 2D (in the image) and in 3D (in the scene).

The classification of moving regions step consists in labelling the moving regions into 8 semantic classes of objects: 4 classes for people (person, occluded person, group, and crowd) and 4 classes for other objects (metro train, scene object, noise and unknown). An occluded person is for example a person partially occluded behind a pillar. A scene object is for instance a seat or a trash. The moving regions can be merged to improve the matching with a class. For instance, two moving regions corresponding to the head and the body of a person can be merged to form a new moving region belonging to the class person.

3. GROUP TRACKING

3.1 Tracking Approach

A real group of people is defined as a set of persons who are close to each other. We detect real groups of people thanks to a structure that we call a group. This group structure is a set of moving regions characterised by four particularities: