Multi-object Motion Pattern Classification for Visual Surveillance and Sports Video Retrieval

Akira Hayashi, Ryuji Nakashima, Toshihiko Kanbara and Nobuo Suematsu
Faculty of Information Sciences, Hiroshima City University
3-4-1 Ozuka-higashi, Asaminami-ku, Hiroshima, 731-3194 Japan
akira@im.hiroshima-cu.ac.jp

Abstract

This paper presents a method to classify scenes based on motion information. While they use object trajectories or optical flow field as motion information in previous work, we use the instantaneous motions of multiple objects in each image. In order to deal with variable number of objects in a scene, we use moment statistics as features. Our approach is based on clustering, a form of unsupervised learning, and needs little human intervention. Furthermore, the probabilistic model based clustering makes it easy to detect scenes with novel patterns.

1 Introduction

Motion information is important for scene activity recognition in visual surveillance [8, 6, 1, 10, 15] and for content based video retrieval [3, 12, 17, 7, 2]. This paper presents a method to classify scenes based on motion information.

In previous work, they use as motion information :

- trajectories of each object [3, 8, 2, 15]
- positional relationship of two objects, and its temporal change [17, 6, 10]
- motion of each pixel (i.e. optical flow field) [12, 7, 1]

We use as motion information the positions and the velocities of multiple objects in each image. Which motion information to use depends on the application. We suppose that the motion information we use is suitable for recognizing the situation in a total scene when there are many moving objects such as people and vehicles in the scene.

Since the number of objects in a scene changes with time, we cannot use object positions and velocities themselves as features for the scene. Instead of them, we use moment statistics for the distribution of the moving objects in position and velocity product space.

The proposed method consists of two phases, the learning phase and the recognition phase. In the learning phase, scenes in the learning data are clustered in the feature space. Suppose the scenes in the learning data correspond to usual situations at the observation site. Then the clustering result is the classification of the scenes for usual situations. In the recognition phase, a newly observed scene is classified according to which cluster the scene belongs. In addition, scenes which do not belong to any of the clusters can be detected as novel scenes.

Generally speaking, (1) which scenes correspond to usual situations, (2) which scenes correspond to novel situations, and (3) how scenes for the usual situations are classified, all of these depend on the application (e.g. the target, the objective, and the method of the surveillance, or the category of the video library, and the objective of the retrieval). But, in our machine learning approach, much of the application dependent knowledge can be learned directly from data through clustering. Hence the proposed method is very much independent from the application. Furthermore, clustering is a form of unsupervised learning and does not need us to label the learning data as usual or novel.

We explain the input data in Sec. 2. We then explain the learning phase and the recognition phase of the proposed method in Sec. 3 and Sec. 4 respectively. In Sec. 5, the result of the experiment using three examples will be discussed. We summarize our presentation in Sec. 6.

2 Input Data

We fix a video camera at an observation site and film continuously. We assume a detection and tracking system for all the objects such as people or vehicles which appear in such video scenes (Interested readers are referred to review papers on tracking [5, 11]. There was a W/S on the subject recently [16].)

At each instance of time t, the tracking system provides