Automated Visual Surveillance Using Hidden Markov Models

Vinod Nair James J. Clark

Centre for Intelligent Machines
McGill University
Montreal, PQ H3A 2A7
{vnair, clark}@cim.mcgill.ca

Abstract

This paper describes an automated visual surveillance system that detects suspicious human activity in a scene. The system is designed to: 1) detect and track people in the scene, 2) recognize the "normal" activities in the scene, and 3) detect anomalous activity by finding sufficiently large deviations from the normal activity patterns. The stochastic time-sequence recognition framework of the Hidden Markov Model (HMM) forms the basis of activity recognition and anomaly detection. We have implemented the system to monitor an office corridor in real-time using a Pentium III machine running Windows 2000. The results show that the system correctly classifies examples of normal activities in the corridor and identifies a mock break-in attempt as suspicious activity.

1. Introduction

Automated visual surveillance is becoming an increasingly important area of research in computer vision. CMU's Video Surveillance and Monitoring (VSAM) project [2] and MIT AI Lab's Forest of Sensors project [7] are examples of recent research efforts in the field. Interest has been motivated by commercial applications such as surveillance of airports and office buildings, as well as military ones, such as monitoring the battlefield to automatically collect strategic information. Conventional visual surveillance systems have limitations that make them less than ideal for many applications. For instance, recording the surveillance video on tapes can provide evidence only after a security breach has occurred. The alternative of dedicating a security worker to watch the live video is expensive and prone to human error. Automated visual surveillance overcomes these limitations by detecting suspicious activity as it happens, without human effort.

Our approach to automated visual surveillance is to classify the normal activities using a set of discrete Hidden Markov Models (HMMs), each trained to recognize one activity, and label the unrecognized activities as unusual. In recent years, HMMs have become popular in computer vision as an activity recognition algorithm. They have been used to recognize hand gestures in sign language [6], facial expressions [4], and different tennis strokes [8]. They have also been used in visual surveillance systems for classifying activities in an office room [1], and in a parking lot [1, 3]. A common feature of these applications is the use of HMMs to generate high-level inferences with only relatively coarse, low-level sensory data, such as blob features. This illustrates an important advantage of HMMs – combining coarse, low-level sensory data with the prior knowledge of the data's statistical characteristics learned by HMMs, avoids the need to compute high-level representations of the data using expensive image processing algorithms.

The paper is organized as follows: section 2 describes the architecture of the surveillance system. Section 3 gives the results of the system's performance, and section 4 presents the conclusions of the paper.

2. System's Architecture

2.1. Hardware Component

The surveillance system uses an ordinary netcam to obtain video of the office corridor under surveillance. The netcam has a built-in HTTP server from which the video can be downloaded as hardware-compressed JPEG frames via the internet, as shown in figure 1. The use of netcams makes the system simple and inexpensive, and allows great

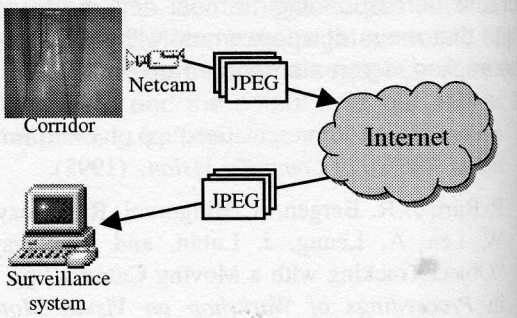

Figure 1. The netcam transmits JPEG frames to the surveillance system via the internet.