**INTRODUCTION**

**T**HE use of visual data such as images and videos forscientific analysis to solve real-world problems has beenincreasing significantly over the past decade. Network cameras are of particular interest as they generate continuous real-time video data with rich and versatile content [31]. Millions of network cameras are deployed every year [24]. The video analysis market is rapidly growing and is estimated to be worth more than $1.2 billion by the year 2017 [6]. A wide variety of applications such as improving public safety [21], aiding emergency response [14], and surveillance [28] may use the large volumes of visual data. The network cameras considered in this paper consist of both indoor and outdoor cameras including traffic cameras, cameras inside shopping malls, and other institutions. These are public cameras providing free access and owned by different organizations. Modifying any configuration settings of the network cameras is impossible.

These applications may require (1) high resolution video data, (2) analysis for long durations, (3) data from multiple cameras, and (4) high frame rates. These requirements represent “big data” problems that require analysis of large amounts of visual data which needs substantial amounts of computational resources. Cloud computing has the potential to meet these resource needs by selecting many cloud instances (i.e., virtual machines, VMs) containing more cores and large memory. Many applications require streaming data from network cameras around the world, for example, studying the traffic pattern of cities, analyzing global fashion trends, and monitoring weather conditions at different regions. The distance between the network camera and cloud instance can affect the performance of the analysis. Hence there is a need to efficiently stream the data from multiple sources at different geographical locations. Cloud vendors offer many types of VM instances: with different number of cores, memory capacities, and geographical locations. The “pay-per-use” pricing model encourages the use of only a few cloud instances with small number of cores and less memory. The computational requirements of the applications may vary depending on the time of the day and the content of the scene being analyzed. The different competing factors mentioned above make resource management a challenging problem. Few studies have been devoted to selecting the most efficient cloud instances to analyze many video streams at low monetary costs. These studies do not consider the effects of the different types and locations of the instances on the overall cost and performance of the analysis.

This paper presents a method called Adaptive Resource Management for Video Analysis in Cloud (**ARMVAC**). ARMVAC determines the configurations (types, locations, and numbers) of cloud instances needed to meet the performance requirements at low costs. We consider Motion JPEG (MJPEG) [4] as the format of video data because most network cameras support MJPEG streaming (some newer network cameras also support H.264). ARMVAC considers the network distances (measured by the round-trip time, RTT) between cameras and cloud instances. ARMVAC models the relationships of the frame rates and CPU utilization on different types of cloud instances.We develop our model through the analysis of three different computer vision methods provided by the OpenCV library [7]: People Detection, Edge Detection, and Color Histogram. We model the problem of selecting cloud instances at low costs as a Variable Size Bin Packing Problem (VSBPP) and use a heuristic algorithm [12] to find a solution. ARMVAC predicts the maximum number of streams from cameras to be analyzed on different cloud instances for the given analysis programs. Our solution can dynamically adapt to the varying resource requirements of analysis programs running for long durations. ARMVAC monitors the utilization of the cloud resources at regular intervals and automatically scales the number of resources based on the utilization and performance requirements. ARMVAC is evaluated using Amazon Elastic Compute Cloud (EC2) instances [2]. Three analysis programs: Motion Estimation, Face Detection, and SIFT feature extraction are used for evaluation. The method can achieve the required frame rates and save up to 62% cost compared with four other cloud resource selection strategies.

This paper makes the following contributions: (1) It is one of the first papers devoted to selecting the cloud configurations for analyzing large (GB) amounts of data from multiple video streams. The sources of the streams are globally distributed. (2) Our method considers both performance requirements and costs, modelling this problem as a bin packing problem and using a heuristic solution. (3) The paper presents a prediction model based on CPU utilization for determining the number of streams that can be analyzed on different types of cloud instances for a given analysis program. (4) We evaluate the solution using Amazon EC2 and demonstrate up to 62% cost reduction compared with four other strategies for selecting cloud instances.