Big Data Analysis

Homework 2

1 The Description of Project

In municipal projects from various countries, there are many video data recorded by cameras, and the utilization of these video data is very low. This project intends to use these video data to perform face recognition and locate the track of where a person appears.

Figure 1: Roadside Camera

This project is based on Spark, Hadoop HDFS, HBase, Kafka and OpenCV to build a video processing and face recognition application system. The system mainly uses the big data processing framework to perform video data processing, face detection, pose optimization, facial

feature extraction and recognition on the video surveillance area.
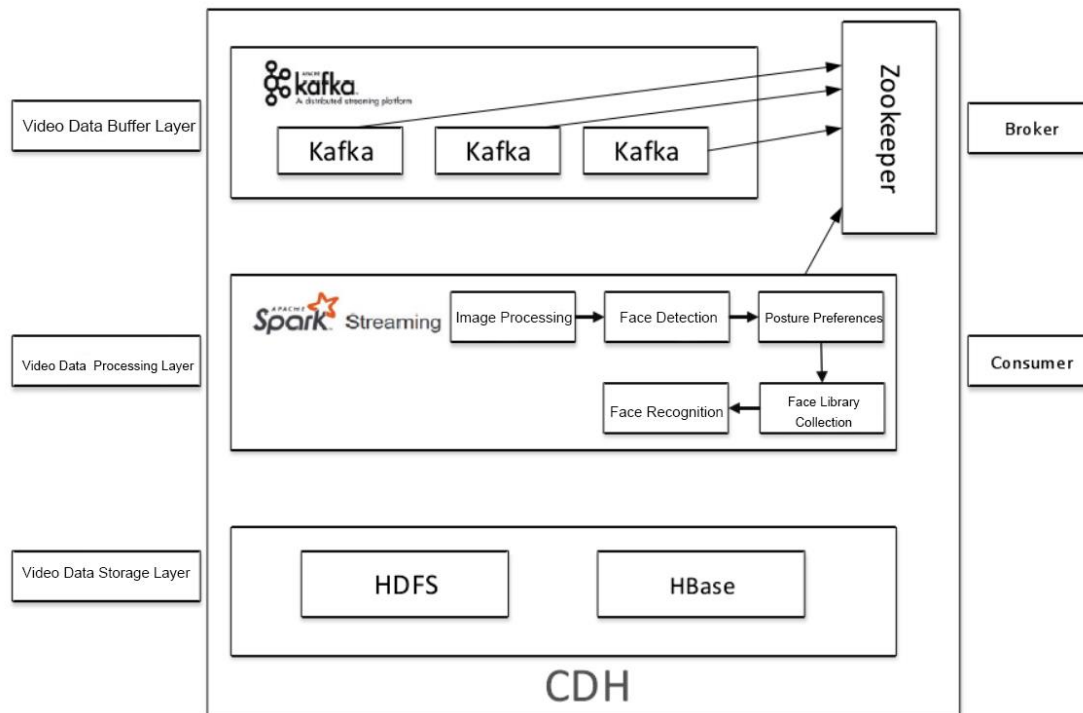
2   The Workflow of the Project

Figure 2: The Workflow of the Project

Video data acquisition layer: The front-end data acquisition device is mainly a high-definition camera, and the video data is obtained through the IP in each acquisition device.

Video data buffer layer: Because the amount of data is huge, data loss is prone to occur. Video data is stored in the buffer memory.

Video data processing layer: This layer can be divided into two parts: video data processing phase and face recognition phase. First obtain the video picture by obtaining the key frames of the video, and then perform face detection, posture estimation and optimization, face feature

extraction, face database collection and face recognition. Lastly, we use the location data to get the track of the person.
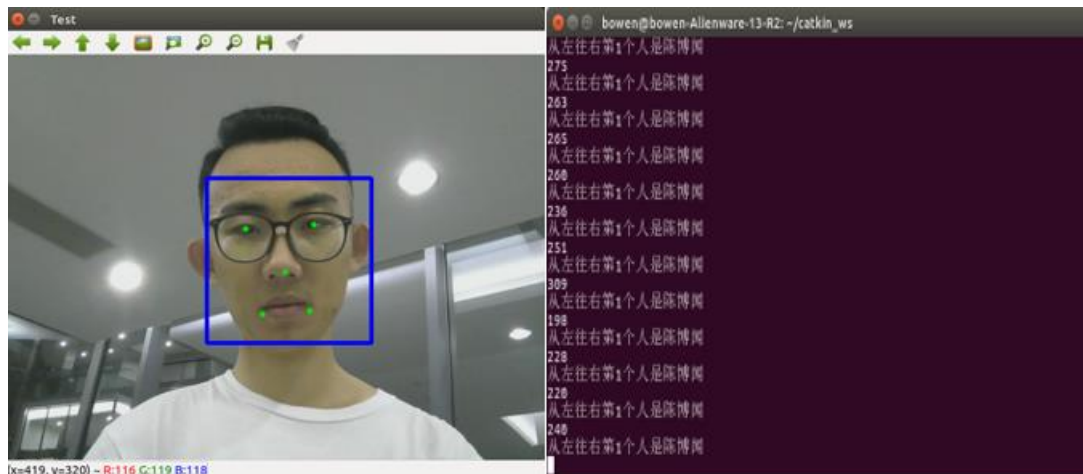

Figure 3: Face Recognition

3   Data Storage

Video data storage layer: Hadoop HDFS and HBase are used as the system storage bottom layer. HDFS is used to store video data, and HBase is used to store the feature description of the processed results, location data and the face pictures.

HDFS is fault-tolerant and is designed to be deployed on low-cost hardware. And it provides high throughput to access application data, suitable for those applications with large data sets. HDFS relaxes the requirements on POSIX and it can achieve streaming access to data in the file system. Therefore, we can use it to store video data.

HBase runs on the HDFS file system and provides Hadoop with services similar to BigTable scale. Additionally, it can provide extremely high fault tolerance for sparse files. Therefore, we can use it to store the

feature description of the processed results, location data and the face

pictures.