Staff identification from a video clip is performed in two stages: footfall detection and nametag extraction. In the footfall detection stage, the RAPiD-T GitHub project was cloned and utilized to detect the footfall in the video clip. RAPiD-T is a combination of the RAPiD algorithm and state-of-the-art (SOTA) object detection algorithms specifically designed for side-view regular cameras, including Robust and Efficient Post-Processing (REPP) and Flow-Guided Feature Aggregation (FGFA). In the nametag extraction step, various image processing techniques are applied to identify the footfall of staff, specifically those who are wearing nametags. The project is being developed using the Python programming language in the Colab environment.

**Footfall Detection**

In the footfall detection stage, the following techniques are applied:

RAPiD Algorithm:

* RAPiD is an algorithm designed for detecting people in overhead fisheye images.
* It addresses challenges posed by fisheye camera images such as distortion, non-uniform resolution, and rotation.
* The algorithm involves preprocessing, rotation-aware CNN detection, and bounding box refinement.
* Preprocessing rectifies the fisheye image to a perspective view to reduce image distortion.
* A deep neural network is used to classify image patches as person or non-person.
* Bounding box refinement is performed using DPM to improve the accuracy of bounding box prediction.

REPP Post Processing:

* REPP (Robust and Efficient Post-Processing) is chosen as the post-processing method for the footfall detection stage.
* REPP is used as the spatio-temporal extension of the RAPiD algorithm, designed to link bounding boxes in consecutive frames.
* It generates object tubelets to smooth out the location and size of bounding boxes.

**Nametag Extraction**

In the nametag extraction stage, the following techniques are applied:

Mask Creation:

* Based on the bounding boxes obtained from footfall detection, a mask is created.
* The mask is applied to the respective frame, masking out everything except the region of interest containing people potentially wearing nametags.

Color Thresholding:

* Color thresholding is applied to the masked region in the HSV color space.
* This isolates colors that are representative of the nametag.
* A binary mask is created, highlighting the nametag colors and suppressing other colors.

Contour Detection and Filtration:

* Contour detection is performed on the binary mask obtained from color thresholding.
* Filtration criteria are applied to identify relevant contours likely to represent nametags.
* Filtering criteria include contour area, aspect ratio, size constraints, and shape.

Histogram Comparison:

* Histogram comparison is employed to compare the histograms of the remaining contours with a reference histogram.
* The reference histogram is obtained from a captured nametag in one of the frames.
* Correlation is used to measure the similarity between histograms.
* The histograms of each region of interest are computed using the S and V channels of the HSV color space.

Based on the remaining contours, regions containing nametags are identified. The respective people in the frame are surrounded by bounding boxes, and a video is created.