Introduction

• The aim of the project is to develop a method to detect whether people are present in a video clip recorded using any mobile device.
• There are several approaches to this problem; different algorithms produce varying results depending on the distance of a person from the recording camera.
• Three methods have been implemented here:
  1. Viola Jones framework for face detection.
  2. Viola Jones framework for upper body omega shape detection.
  3. Histogram of oriented gradients features with support vector machine classifier.

Viola Jones Framework for Object Detection

• Faces and upper body omega shapes are detected by the standard procedure of feature extraction and region classification.
• Haar-like features are extracted very fast by using the concept of integral images.

Haar-like Features

\[
\begin{align*}
(a) & \text{ 3-rectangle feature} \\
(b) & \text{ 2-rectangle feature} \\
(c) & \text{ 1-rectangle feature} \\
(d) & \text{ 4-rectangle feature}
\end{align*}
\]

• Feature = \(\text{sum(pixels in black region)} - \text{sum(pixels in white region)}\)
• A cascade of AdaBoost classifiers uses the features to quickly reject irrelevant part of the image and to identify the relevant part, viz. face.
• This framework has been used for:
  1. Face Detection.
  2. Upper Body Omega Shape Detection.

Related Work

• Face Detection algorithms are being used for several applications:
  • Skin color based detection.
  • Facial features (eyes, nose, etc.) based detection.
• Video processing techniques for person detection:
  • Background subtraction to detect moving bodies.
  • Human motion tracking.
• Face detection is extensively used in cameras nowadays.

Experiments and Evaluation

• Experiments performed on videos captured at different locations around Los Angeles such as the Universal Studios, a beach, the USC campus, downtown Los Angeles using the MediaQ app.

Viola Jones: Face Detection

HOG: Standing Posture Detection

• A problem with the Viola Jones framework with for upper body omega shape detection was that it resulted in many false positives.
• Remedy: using a HOG-based SVM classifier in cascade with Viola Jones framework.

Histogram of Oriented Gradients

• A lot of characteristic information is stored in the structure of standing posture of the human body.
• Each detection window of size 64x128 is divided into a number of fixed size blocks and each block is further divided into cells.
• A histogram of gradient values is computed for each cell with 9 orientation bins between 0 and 180 degrees.
• The gradients are calculated using the basic 1-D mask as shown below.
  \([-1 \ 0 \ 1]\)
• A linear SVM is trained using HOG descriptors obtained from images of standing people and it is then used for full body detection.

Conclusion and Scope for Further Work

• The three algorithms perform fairly well when they operate independently.
• Videos having no human presence are detected with high accuracy.
• The performance of the system could be improved with a hybrid approach to obtain an accurate count of people per frame in a video.
• Future work could include utilizing:
  i. Spatial Aspect: Eliminating multiple detections from different algorithms to give an accurate estimate of number of persons in a frame.
  ii. Temporal Aspect: time saving by searching for human presence in a frame in the neighborhood of windows detected in the previous frame.