Table of Contents

1 Introduction

2 Methodology
   • Feature Extraction
   • Classification
   • Object localization

3 Experiments and Results

4 Scope of Improvements
Outline

1 Introduction

2 Methodology
   - Feature Extraction
   - Classification
   - Object localization

3 Experiments and Results

4 Scope of Improvements
Introduction

Aim

- **Detection** and **classification** of objects in surveillance video into various categories.
- **Categories** - Car, Person, Motorcycle, Bicycle, Rickshaw, Autorickshaw

**Figure:** Detection and Classification of objects

**Motivation** - Recognition, Tracking, Security, etc.
Outline

1 Introduction

2 Methodology
   - Feature Extraction
   - Classification
   - Object localization

3 Experiments and Results

4 Scope of Improvements
Outline

1. Introduction

2. Methodology
   - Feature Extraction
   - Classification
   - Object localization

3. Experiments and Results

4. Scope of Improvements
To obtain the image dataset required for classification, we extracted frames from the provided video data. Frames were taken after every 20 frames of each video. The images in the dataset correspond to the labelled bounding boxes from these frames.

Figure: Red Boxes Images used as dataset
Feature Extraction - Feature Vector

Bag of visual words[1]

- Computed $\sim 300$ SIFT points per image in the dataset.
- Used mini-batch K-means - 700 clusters of siftpoints.
- Voting of siftpoints of each image in clusters give vector of dimension 700.
Feature Extraction - Feature Vector

Bag of visual words[1]
- Computed \( \sim 300 \) SIFT points per image in the dataset.
- Used mini-batch K-means - 700 clusters of siftpoints.
- Voting of siftpoints of each image in clusters give vector of dimension 700.

HOG features[2]
- Images in the dataset are resized to 128x128.
- Obtained HOG feature vector of dimension 15876 for each image.
Feature Extraction - Feature Vector

Bag of visual words[1]
- Computed $\sim$ 300 SIFT points per image in the dataset.
- Used mini-batch K-means - 700 clusters of siftpoints.
- Voting of siftpoints of each image in clusters give vector of dimension 700.

HOG features[2]
- Images in the dataset are resized to 128x128.
- Obtained HOG feature vector of dimension 15876 for each image.

Convolution Neural Network features
- Used VGG 16 Model[3].
- Output of FC-4096 layer as feature vector.
Outline

1. Introduction

2. Methodology
   - Feature Extraction
   - Classification
   - Object localization

3. Experiments and Results

4. Scope of Improvements
After obtaining feature vectors of images in the dataset, classifiers are trained given the category of each image.

**Classifiers Used:**
- Support Vector Machines
  - Linear Kernel
  - Gaussian Kernel
- Random Forest Classifier
- Decision Tree
1. Introduction

2. Methodology
   - Feature Extraction
   - Classification
   - Object localization

3. Experiments and Results

4. Scope of Improvements
Object localization

Objects of interest are extracted from the test videos. Three techniques are implemented for the task.
Object localization

Objects of interest are extracted from the test videos. Three techniques are implemented for the task.

- **Using Background Subtraction**
  - Using Mixture Of Gaussions (MOG)
Object localization

Objects of interest are extracted from the test videos. Three techniques are implemented for the task.

- **Using Background Subtraction**
  - Using Mixture Of Gaussions (MOG)
- **Measuring Optical Flow**
Objects of interest are extracted from the test videos. Three techniques are implemented for the task.

- **Using Background Subtraction**
  - Using Mixture Of Gaussions (MOG)

- **Measuring Optical Flow**

- **Sliding Window Method**
Object localization

Objects of interest are extracted from the test videos. Three techniques are implemented for the task.

- **Using Background Subtraction**
  - Using Mixture Of Gaussions (MOG)
- **Measuring Optical Flow**
- **Sliding Window Method**

![Figure: Contours are drawn on right image](image-url)
Object localization

Objects of interest are extracted from the test videos. Three techniques are implemented for the task.

- **Using Background Subtraction**
  - Using Mixture Of Gaussions (MOG)
- **Measuring Optical Flow**
- **Sliding Window Method**

**Figure:** Contours are drawn on right image
Then, Learned Classifier is used to classify extracted image.
1 Introduction

2 Methodology
   - Feature Extraction
   - Classification
   - Object localization

3 Experiments and Results

4 Scope of Improvements
Experiments and Results

Dataset for classifier learning

- Labelled dataset was noisy - separated good images for dataset creation.
- ~1600 training images and ~450 test images.
- Around 300 training images of each class used for training.

Accuracy Obtained

<table>
<thead>
<tr>
<th>Classifier/features</th>
<th>BOVW</th>
<th>HOG</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM (linear)</td>
<td>0.8535</td>
<td>0.9129</td>
<td>0.9873</td>
</tr>
<tr>
<td>SVM (gaussian)</td>
<td>0.879</td>
<td>0.756</td>
<td>0.9809</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.8322</td>
<td>0.9023</td>
<td>0.9745</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.6475</td>
<td>0.6072</td>
<td>0.8917</td>
</tr>
</tbody>
</table>
Experiments and Results

Dataset for classifier learning

- Labelled dataset was noisy - separated good images for dataset creation.
- \( \sim 1600 \) training images and \( \sim 450 \) test images.
- Around 300 training images of each class used for training.

Accuracy Obtained

<table>
<thead>
<tr>
<th>Classifier/features</th>
<th>BOVW</th>
<th>HOG</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM(linear)</td>
<td>0.8535</td>
<td>0.9129</td>
<td>0.9873</td>
</tr>
<tr>
<td>SVM(gaussian)</td>
<td>0.879</td>
<td>0.756</td>
<td>0.9809</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.8322</td>
<td>0.9023</td>
<td>0.9745</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.6475</td>
<td>0.6072</td>
<td>0.8917</td>
</tr>
</tbody>
</table>
Outline

1. Introduction

2. Methodology
   - Feature Extraction
   - Classification
   - Object localization

3. Experiments and Results

4. Scope of Improvements
### Scope of Improvements

**Dataset**

Dataset and Labels are noisy, improvement of dataset are likely to give improve results.
Scope of Improvements

**Dataset**  Dataset and Labels are noisy, improvement of dataset are likely to give improve results.

**Localization**  BGS, Optical flow, Sliding window do not work good for localization. Learning method can be used for localization.
Scope of Improvements

**Dataset** Dataset and Labels are noisy, improvement of dataset are likely to give improve results.

**Localization** BGS, Optical flow, Sliding window do not work good for localization. Learning method can be used for localization.

**CNN** Though the accuracy of classifier with CNN features is very good, it require 1-2 second each frame for the classification. For real time classification, Fast R-CNN[4] can be used.
Object recognition from local scale-invariant features
Proceedings of the International Conference on Computer Vision. pp. 1150-1157

Histograms of oriented gradients for human detection
Dalal, Navneet, and Bill Triggs.

Very Deep Convolutional Networks for Large-Scale Image Recognition
K. Simonyan, A. Zisserman.
arXiv:1409.1556
Fast R-CNN
Ross B. Girshick.
Proceedings of the IEEE International Conference on Computer Vision. 2015.

Libraries Used
OpenCV(i/o, bgs, optical flow, sift), sklearn(k-means, classifiers), sklearn-image(hog).