1. Introduction

Computers have been used intensively in our daily lives. In the past, people used them to speed up complex calculation. Moreover, computers nowadays are not only the big calculators, but it also can simulate human perspectives. Computer scientists have tried to extend computer capability in order to allow smart machine to substitute human in some certain jobs such as dangerous or poisonous ones. They have added artificial intelligent to computer to make it seems to have perspectives as human. Among enormous applications of computer, computer vision is the subject which draws most attention of computer scientists.

Our daily lives is filled of millions of objects ranging from big ones such as human, car, bicycle,… to tiny ones like cells. And the task of recognition and classification each object to its catalogue is the fundamental task for any intelligent based system. The difficulty is that a given class has a huge intra class variation. For example, human is usually thought as an object consists of up-right shape, two legs, two hands, and an omega shaped head. However, in reality, human probably appear in diverse shape. For example, people who sit down, stand up, lie down, or play sport have totally different shape. In addition, illumination, points of view are also the significant factors affect to recognition and detection process.

Thus, recently, the goal of researchers working in computer vision and intelligent based machine is to invent algorithms or facilities in order to allow computer has the ability to see and analysis a given images or videos. And one of the primary tasks is the detection and catalogue objects in images. Such ability allows us to have numerous applications such as human computer interaction, robotics, smart autonomous vehicle as well as image retrieval.

In this chapter …

* 1. Goal and Applications
     1. The goal

The main target of this thesis is to build up an automatic system which is possible to detect and localize pedestrians in static image. For more specific, it is the issue of create object detection from the view of point of computer, in which detector scan all the given image and bound the box around object if it appears in image. We use the approach which utilize robust extraction algorithm to extract a region of an image, and then use a classifier to decide whether this region contains pedestrian or not. In this cope of thesis, we just concentrate on how to encoding image regions into feature vectors which is robust on illumination, slight change and osculation.

Unlike matching one word with another word in which we can easily see they are identical or different, but matching object with object (for example, human) is the totally different matter. Natural object such as human, cat, dog and man-made object such as car, bicycle have diverse of shape, so it is difficult for computer to distinguish two catalogues. In this thesis, we use an approach which does not make strong assumption on context. For example, the context of car in cartoon is wider perspective than the car in show room. So, if we heavily depend on context of the car, we will miss the car in other view of point, such as car in painting or cartoon. Overall, the goal is to build a detector which can detect general object in wide perspective.

* Input: arbitrary image.
* Output: boundary box which contains pedestrian if image has that one.
  + 1. Applications

Robust extraction algorithm is not only useful in finding pedestrians in images, it also can be used to extract characteristic of any object. So, we can use this descriptor as a core in system of analyzing and cataloguing images in album. We obviously see that the advent of digital camera has allowed people to take photograph more easily. In 2-3 years, one personal digital camera can take as many as 10,000 photos, and which is impossible for human to manually search and locate these photos in short time. Consequently, Intelligent Management Software which can automatically add tags to these images to facilitate search is dispensable.

Moreover, person detectors are also being employed for detect pedestrians in smart cars. For instance, a warning message will appear in windshield to arouse drivers whenever the car tends to hit pedestrians or obstacle. Another application in smart car system is that cameras can detect the behaviors and consciousness of drivers in order to execute some proper assistance.

Information detected in multi-cameras will be fused together; and with the training knowledge in system, detectors will make reasoning decision to whether take a certain action or not. However, there is almost no detector which is good performance can execute in real time. For example, with the limited capacity of processing unit of portable devices, it is really hard for them to use good performance detectors in real time. Fortunately, in recent years, by the breakthrough in chip processing, and associating with some good detectors, building the software for smart cars is the subject that draws a lot of attention of researchers.

* 1. Challenge

The most difficulty of building an object detector is the diverse of variation in images. These following factors effect on object detector are:

* Image is just a matter of pixel, and it lacks of motion knowledge like in video.
* Object in image suppress 3-D information and depend on viewpoint of camera as well as the scale.
* AS mentioned above, most natural object classes have huge variation in intra-class. Although two instances belong to one object class, they probably appear different on account for illumination, viewpoint, and shape distorting.
* Background information is also the vital key to prevent us from building robust detectors. Background clutter varies from image to image. For example, images can be taken from indoor, outdoor, and under diverse natural factors such as illumination, viewpoint. So, the desirable detectors have to have the ability of distinguishing object in complex background.
* In image, color and illumination of objects in one class probably varies considerable. Let’s think of a photograph taken in day with direct sunlight and shadows versus one taken in night with dim light, you easily see how the big gap they have. So, the robust detector must have capacity of resisting of changing color and illumination in object.
* Partial occlusion is an inevitable in real images. In this situation, just only a part of object can be visible. That is the reason why creating good performance detectors is very difficult.
  1. Some background of object detection

1.3.1 Image filter

The concept “image filter” has widely used in computer vision. It is use in edge detection, corner detection, blob detection, and noise elimination. A mask which is a matrix or a vector is used to convolve with an image. The most popular mask is Gaussian mask

1.3.2 Gradient

1.3.3 Non-maxima suppression

1.3.4 Linear SVM

* 1. Overview of our approach

In give image, we use sliding window to densely scan at all position at different scales. At each position, we get its score, and we decide this window contains object or non-object via classifier. This method is purely based on statistic approach which disregards the fore-given context of any object class. When extracting region containing object, we assume that there are some invariants which are not change dramatically within one type of object. These invariants become the main characteristics for classifier to distinguish this object class with other object class. So, by extracting invariants of object or non-object, we can represent them in high dimensional vector. And we assume that it is possible to build up a hyper-plane which separates object, non-object point as far as possible.

We just focus on method of represent robust features in order to robust from slightly changes in shape, illumination and scale. The classifier used in this thesis is Linear Support Vector Machine (stand for SVM). Recently, SVM have widely used in machine learning. And in computer vision, it is intensively used in learning process. We use SVM because it is simple, runs fast, and has good performance.

For more specific, in extracting feature process, we use locally normalized Histogram of Oriented Gradients (HOG) as a descriptor. HOG is computed from gradients of image and has the characteristic that robust to (1) small changes in image contour locations and directions, (2) significant change in image illumination and color, (3) remaining as discriminative and reparable as possible. We use weighted histograms gradient orientations over spatial neighborhood to calculate HOG features. Before calculate histogram of gradients, we do some pre-process to eliminate the effects of illumination and color changes. So, the histogram of oriented gradients has information of the contour of the object.

Once we densely scan image, we will get a bulk of windows at level classifier which means that each window is now represented as high dimensional feature vector. Note that we scan all position at multiple scales, so there are probably some windows overlap each other. After that, we suppress all window whose score below the threshold, and keep and positive windows (exceed the threshold). Because HOG is robust to slight changes in shape and contour, it is possible to have many positive windows contain same object. To resolve this problem, we fuse all positive ones and use non-maxima suppression to find only one window most likely contains object of a class. In this thesis, Mean Shift is used as a suppression algorithm in this process.

After densely scanning image, we filter windows which exceed the threshold. We assume that our extract algorithm is robust which means that the score of window is still large (not maximum) even if this window slightly off center of object. And then, we use non maximal suppression (such as mean shift) to find the mode.

In process of extract a slide window to feature vector, we use robust extraction algorithm to avoid the noise effects of background. And HOG (histogram of gradient) seems to be the best choice because it is invariance with variety of background. In realistic, HOG is declared as the state of the art in extract feature vector to distinguish object or non-object.

INRIA pedestrian dataset is intensively used in training phase and testing phase. In training phase, there are 2416 positive windows and 1220 negative images; and there are 1126 positive windows and 453 negative images.



Pedestrian/ non-pedestrian classifier

Figure 1: Overview of detection method

1. At each point in image, we densely scan with multi-scale. (b) Sliding window is detected and extracted to feature vector which is input to pedestrian/non-pedestrian classifier.

**Overview of extraction algorithm**

* Normalize gamma, using equalizeHist function in OpenCv in each channel color.
* At each color channel, compute gradient of each pixel using centered mask, the info of gradient is the weight and orientation. And at each pixel, we choose color with greatest magnitude of gradient.
* Divide window into many equal squared cells.
* In each pixel, we calculate how it contributes to histogram of cell which contain it, as well as the neighborhood cells. (spatial and orientation histogram). (See figure 2)
* In the window, blocks are created by group of cells. Blocks can overlap each other.
* Concatenate histograms of blocks to form feature vector of window.

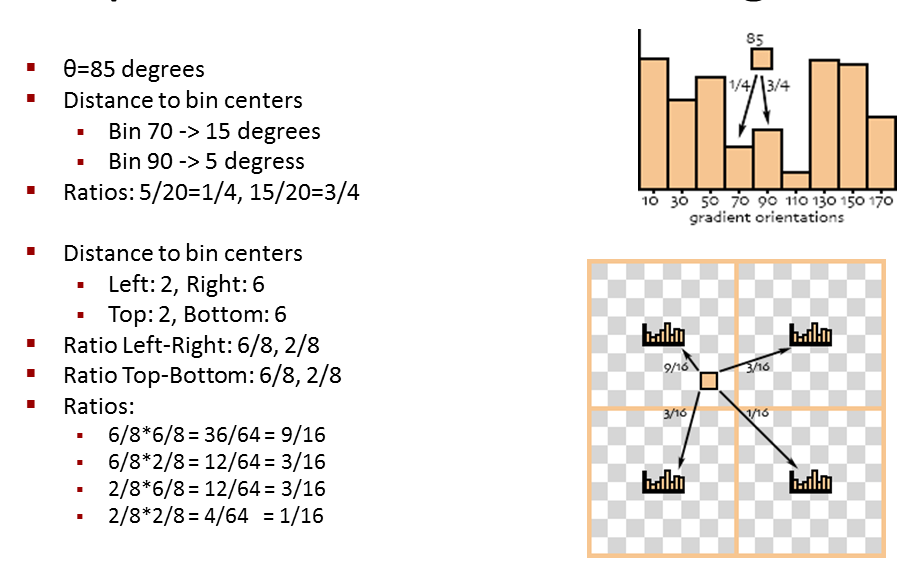


Figure 2: Example of Spatial & orientation histogram

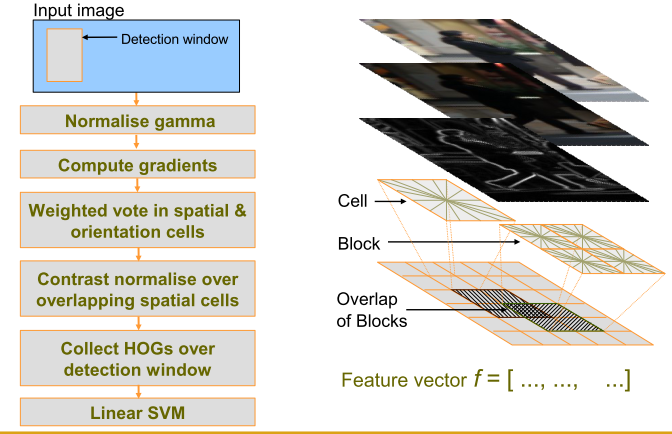


Figure 3: Overview extraction algorithm Histogram of Oriented Gradient

* 1. Overview of result

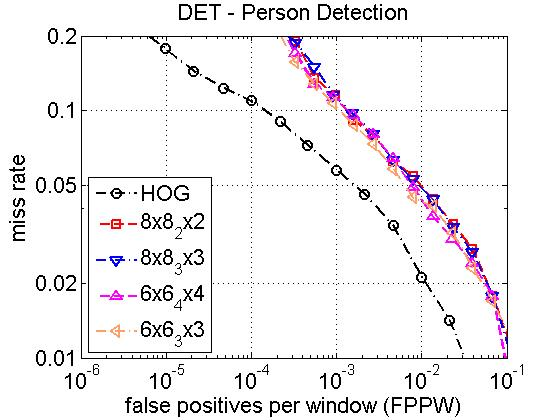


Figure 4: Performance of Dalal and us. Dalal’s curve is the first one (black).

* 1. Outline

1. Related work
2. Overview of Detection Methodology and results
3. HOG
4. Some modified HOG
5. Multi scale object localization
6. Dataset
7. Conclusion
8. Future work