International Journal of Thesis Projects and Dissertations (IJTPD)
Vol. 2, Issue 2, PP: (15-21), Month: April - June 2014, Available At: www.researchpublish.com

Car Detection in UAV Images

Hriday Bavle

Department of Space Science and Technology

Amity University, Noida,India

Abstract: **Object detection is a very important problem in the field of computer vision. Object detection is the process of finding instances of objects such as faces, bicycles, and buildings in images or videos. Object detection algorithms generally use extracted features and learning algorithms to recognize instances of an object category. It is commonly used in applications such as security, image retrieval, surveillance, and automated vehicle parking systems. This paper tries to provide an extension to this problem,by detecting objects in the images obtained from the UAV. The object focused here is only a single object i.e. car. This paper uses bag of words model with scale invariant feature transform. The video obtained from the UAV is analyzed frame by frame by the program and the number of cars in each frame of the video arefirst detected and the detected cars are then counted and displayed in the end.**

Keywords: **Car Detection, Feature Extraction, Scale Invariant Feature Transform (SIFT), Support Vector Machine (SVM), Unmanned Aerial Vehicle (UAV).**

I. INTRODUCTION

There are several techniques used in computer vision for detection of objects namely feature based object detection, viola jones object detection, image segmentation and blob analysis etc. In this paper the car is detected using feature based object detection. Feature based object detection has several algorithms like the scale invariant feature transform, speeded up robust features etc. to compute the interest point descriptors and match them. This paper focuses on SIFT algorithm to calculate the interest points and compute their descriptors in the images containing several objects including car, these image descriptors are first used by the bag of words model to form visual words in the several car images using clustering algorithms.These visual words are then used to represent certain car and non-car images (training set) by computing the bag of words descriptor. A classifier is then trained for the training. This trained classifier is then tested on a given query images to detect the presence or absence of car.

II. Bag of Words Model with SIFT

The bag of words model is derived from natural language processing where the occurrence of words is more important than the order of the words. For example consider a problem of classifying conference papers, first step is creation of a dictionary of several words. In the second step several papers (training set) are analysed by the classifier and in the third step while testing the classifier, random conference papers are classified according to the frequency of words in the paper.

Similar assumptions are made in image processing, where images can be classified according to the occurrence count of the visual words and not their spatial orientation.

Thus object detection using bag of words model with SIFT can be achieved in 4 steps namely

1. Feature Extraction using SIFT

2. Quantization

3. Bag of Words Image Representation

4. Image Classification

1. Feature Extraction using SIFT

The SIFT algorithm was proposed by David Lowe (2004), in this algorithm the SIFT keypoint and descriptors are computed. These SIFT descriptors are invariant to following changes,

- Scale
- Rotation
- Illumination
- Viewpoint

Hence they can be used for accurate object detection using feature matching.

Steps in the extracting the SIFT descriptors are:

1.1 Construction of Scale Space in SIFT

1.2 Difference of Gaussian

1.3 Keypoint Localization

1.4 Generation of SIFT Feature Vector

1.1 Construction of Scale Space in SIFT

Scale Space in SIFT is represented by octaves (k) and scale (σ), were each octave consists of increasing number of scales. The author of SIFT suggests the 4 octaves and 5 scales in each octave.

1.2 Difference of Gaussian

Difference of Gaussian is used to compute the difference between two consecutive scales in an octave. This DOG computes the edges and corners in an image which can serve as potential key points.

1.3 Keypoint Localization

The key point in a particular scale can be computed if it a maxima/minima between its 8 neighbours in the same scale and other 18 neighbours in the scales above and below it. This can be done using a Hessian matrix.

1.4 Generation of SIFT Feature Vector

Generation of a SIFT feature vector is generation of a finger for the keypoint which is unique for a given keypoint.

The space near a keypoint is divided into a 16×16 window, this window is then further divide into 16 4×4 windows, in which the orientation and magnitude of pixels are computed and put in an 8 bin histogram. This being done for all 4×4 it gives a total of 4×4×8 i.e. 128 digit feature vector which is unique for each keypoint and which can be used for matching.

2. Quantization

Quantization is the process of formation of the visual words dictionary. In this process the SIFT feature vectors which are computed for an images in the previous step are clustered using unsupervised learning (k- means clustering) to obtain visual words, which thus forms a dictionary.

3. Image Representation

For any given image, the bag of words representation can be given as a histogram of frequency counts of image features which is obtained when compared to the visual words dictionary. The histogram thus obtained is called as a bag of words descriptor.

4. Image Classification

Given a set a positive and negative images it is necessary to classify these images using a decision boundary.

This can be achieved in two steps namely:
4.1 Training
4.2 Testing

4.1 Training

In the first step the bag of words descriptors for positive and negative images are first labelled. These labelled images are then used to train a classifier like support vector machine.

The support vector machine thus outputs an optimal hyperplane using the support vectors, which is then used to classify unknown images.

4.2 Testing

In the second step the trained classifier is then tested on several random images for presence or absence of a particular object in those images.

III. Car Detection

The above algorithm is written is OpenCV. The car detection process used in this paper can be broken down into the following steps

1. Obtain the Bag of Words Visual Dictionary

- Select a large set of car images
- Extract the SIFT feature points of all the images in the set and obtain the SIFT descriptor for each feature point extracted from the image
- Cluster the set feature descriptors for the amount of bag that have been defined and train the bags with clustered feature descriptors.
- Obtain a visual vocabulary

2. Obtain the Bag of WordsDescriptors for Training Image Set

- Extract a set of feature points from a training set of images i.e. car (positive) as well as non-car (negative) images
- Obtain SIFT descriptor for each feature point
- Match the all the feature descriptors of the training set with the vocabulary
- Build the histogram and label it for positive as well as negative images
- Train the histogram for several positive and negative samples using support vector machine

3. Obtain the Bag of Words Descriptors for a Query Images/Video frames

- Extract the set of feature points from a query image/video frame
- Obtain the SIFT descriptor for each feature point
- Match all the feature descriptors of the query image/video frame
- Build a histogram
- Test the histogram of the query image using the trained classifier for the presence or absence of car
- Count the number of cars present in the frames analysed and display the result

IV. Experiments Performed

Several experiments were performed to create the car detector. Firstly a car database of cars was made using 60 car images taken from a UAV. This car database was first used to make the bag of words dictionary.

After the creation of this dictionary several positive images of cars from the database and several negative images of no-cars also taken by the UAV where labelled and the SVM was trained.

This trained SVM was then tested on a desirable frame of a video recorded by the UAV to detect the presence of absence of car.

The working of the code can be shown by the following flow diagram.

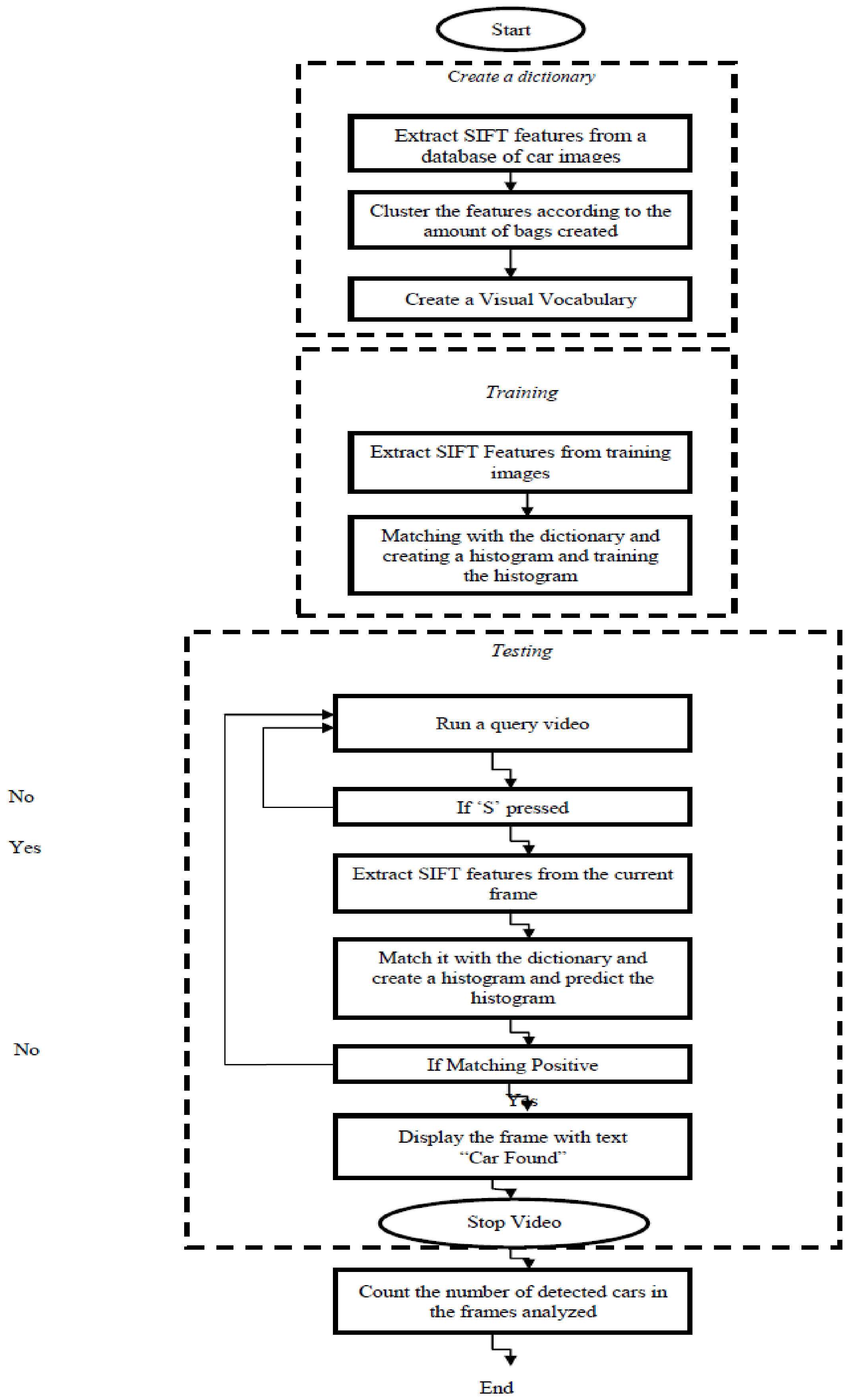

Table 1: Performance of code by changing the number training images

Serial Number	Number of positive images trained	Number of negative images trained	Correct detections	Error percentage
1	549	499	7/10	70%
2	148	104	6/10	60%
3	50	50	8/10	80%
4	14	14	8/10	80%
5	10	10	6/10	60%

From the above table it can be seen that the accuracy of detection of car by the code increases as the number of positive and negative image samples become equal. But the accuracy can be maintained only upto a certain minimum threshold after which the accuracy again decreases. Here this threshold is 10 images.

The below figures show the output of the code,

Fig 1: Detected Car in a Video Frame

Fig 2: No Detected Car in a Video Frame

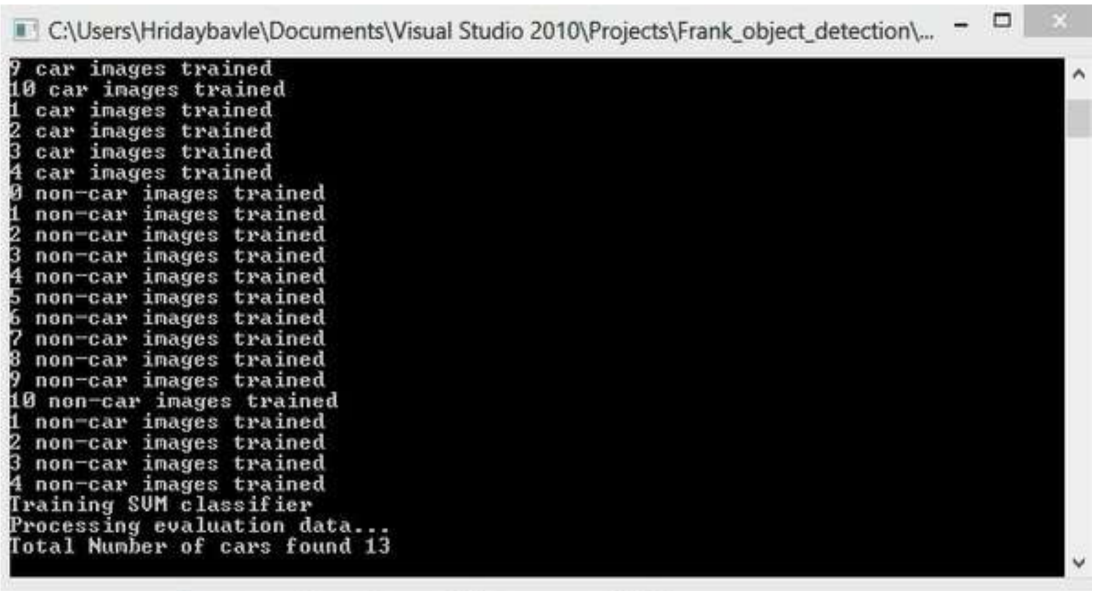

Fig 7: Final Result showing the Total number of counted cars

V. Conclusion

A car detection and counting system using the bag of words model is implemented in this paper. The detection is basically is divided into 3 parts as follows:

1. Creating a dictionary of Bag of Visual Words

2. Training the Classifier

3. Testing the Classifier

This paper uses the Bag of word model to create a dictionary of features of cars which are then matched with other image features to create a histogram.

Since SIFT descriptors are used in this paper, the features extracted from the images are invariant to scale, rotation and illumination and hence it increases the accuracy of matching of features.

This paper then uses the Support vector machine to train the histogram of car features and test it on several query images/video frames.

This system can be used in parking lots where the UAV can be flown and it could detect and count the number of cars present. This system can also be used in a traffic signal or toll booths to count the number of cars that passed through it in a certain period of time. Other applications may include count the number of cars entering a building for a particular time interval.

The future scope of this work could be extended to UAV’s where the detected car could be tracked by the UAV.

Also several objects like pedestrians, bikes, buildings could be detected and classified by the code and these classifications could be fedbackto the UAV. This would create a UAV system capable of detecting and classifying several objects in the scene given below, thus making it artificially intelligent.

REFERENCES

[1] David G.Lowe, “Distinctive Image Featuresfrom Scale-Invariant Keypoints”. International journal of computer vision, Jan 5, 2004.

[2] Inad A. Aljarrah,Ahmed S. Ghorab,Ismail M. Khater“Object Recognition System using Template Matching Based on Signature and Principal Component Analysis”.International Journal of Digital Information and Wireless Communications (IJDIWC), 2012

[3] Herbert Bay, Tinne Tuytelaars2, and Luc Van Gool, “Speeded Up Robust Features”. Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

[4] Gary Bradski and Adrian Kaehler, “Learning OpenCv”. O’Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

[5] ThomasMoranduzzo and FaridMelgani, *“*A SIFT-SVM Method for DetectingCars inUAV Images”.

Department of Information Engineering and Computer Science, University of Trento,

38123 Trento, Italy

[6] Kevin Chang, Todd Swinson,“Using Support Vector Machines to Classify Traffic Signs for an Autonomous

Vehicle”

[7] Richard Szeliski, “Computer Vision:Algorithms and Applications”. September 3, 2010 draft, 2010 Springer

[8] Rafael C. Gonzalez, Richard E. Woods, “Digital Image Processing”. Prentice Hall, 2008

[9] Li Fei-fei, “Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories”, 2004

[10] CordeliaSchmid, Roger Mohr, Christian Bauckhage, “Evaluation of Interest Point Detectors”, 2000